Predicting eucalypt distributions in Tasmania

An application of generalised linear modelling

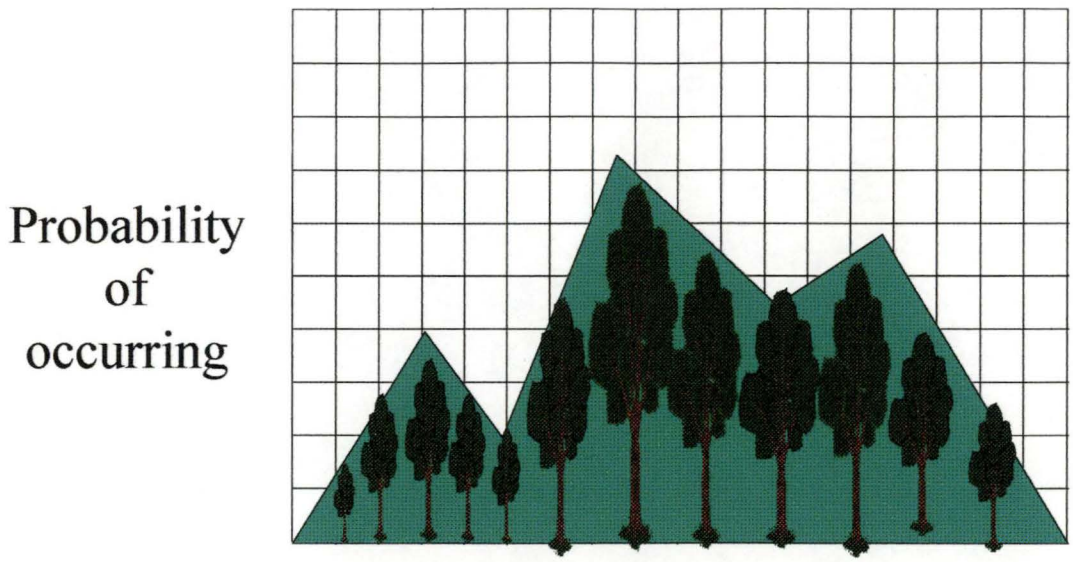
Volume One

Kristen J. Williams B.Sc.(Hons)

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

University of Tasmania
November 1997

Volume One of Two



Picture by Mike Battaglia, October 1997.

Declaration

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material published or written by another person except where due acknowledgment is made in the text of the thesis.

Kristen J Williams
10 November 1997

Access authority

This thesis may be made available for loan and limited copying in accordance with the *Copyright Act 1968*.

Kristen J Williams
10 November 1997

Nomenclature

Vascular plant species nomenclature used in this thesis follows:
Buchanan, A. M. (1995). *A census of the vascular plants of Tasmania and index to the Students Flora of Tasmania*. Tasmanian Herbarium, Hobart, Tasmania.

Abstract

This thesis develops a systematic approach to the routine prediction of *Eucalyptus* species' distributions in Tasmania from compiled ecological data comprising over 15 500 observations. The method of logistic regression, being an application of generalised linear modelling, was used to correlate species' occurrence with environment. Preliminary analyses tested sampling adequacy in terms of ecological variability and species' ranges, and derived environmental indices that could be directly related to plant physiological processes. Subsequent realised niche models were derived for the distribution of *E. globulus* in eastern regions of Tasmania considering biotic and abiotic attributes as predictors and relative dominance as a response in addition to occurrence. Different aspects of the ecology of this species were explored by considering response variables defined by vegetation class or pure and mixed stand occurrences of *E. globulus* and related species from the series *Viminales*. The results of these predictive models were displayed as nested, univariate responses along environmental gradients, representing a direct gradient analysis that facilitated their interpretation in terms of ecological theory and plant physiological processes.

Thesis summary

The objective of this thesis was to develop methods for the routine prediction of *Eucalyptus* species' distribution from an *ad hoc* set of compiled ecological data for presence/absence responses and environmental correlates. These data comprise over 15 500 observations and are typical of that available to land management agencies. The method of logistic regression, being an application of generalised linear modelling, was used to develop predictive models. A systematic approach to analysis was designed to take into account both statistical assumptions and ecological theory.

Preliminary analyses defined sampling requirements in terms of ecological variability and species' ranges, and derived environmental indices that could be directly related to plant physiological processes. Subsequent realised niche models were derived for the distribution of *E. globulus* in eastern regions of Tasmania considering biotic and abiotic attributes as predictors and relative dominance as a response in addition to occurrence. Different aspects of the ecology of this species were explored by considering response variables defined by vegetation class or pure and mixed stand occurrences of *E. globulus* and related species from the series *Viminales*. The results of these predictive models were displayed as nested, univariate responses along environmental gradients, representing a direct gradient analysis that facilitated their interpretation in terms of ecological theory and plant physiological processes.

A novel approach was taken to assessing the representation of forest habitats in the compiled set of ecological data using a simple method of landscape stratification based on environmental factors thought to be correlated with plant ecological responses. A randomised resampling technique was used with the species-area type relationship to estimate the potential number of combinations of environment in a given area for each biogeographic region in Tasmania. This allowed the estimation of a minimum representative sample for different types of forest habitat in each region. Subsequent assessment of sampling adequacy indicated that the ecological data would reasonably represent the ecological relationships for *Eucalyptus* species in mid- to lowland habitats of eastern regions of Tasmania, and that predictions within these regions might be applicable with a reasonable degree of accuracy at broad spatial scales (i.e. a mapping resolution of about 1:500 000 scale).

Predictive models of individual species also require that the sampling domain for the set of absence records that are included with the presence records be defined. There also existed the possibility that species' ecological response functions might be distorted by absence records beyond the environmental range of its presences. Since the distribution of Tasmanian eucalypts had not been previously mapped in a systematic manner, an atlas of the 29 *Eucalyptus* species, comprising over sixty thousand records, was compiled. The range of each species determined with this atlas provided a systematic and objective means of defining the appropriate set of presence and absence records for prediction. The atlas records also provided a context for assessing the geographic and environmental representation of each *Eucalyptus* species in the ecological dataset. Close to half of the 29 species were reasonably represented across two-thirds or more of their potential geographic and environmental ranges and were suitable for use in

prediction. These included both the major ecological keystone species (e.g. *E. amygdalina*, *E. viminalis*) and those species of economic significance (e.g. *E. obliqua*, *E. globulus*).

The thesis subsequently explored the suitability of a range of environmental measures for species' distribution prediction using the criteria of predictive power and interpretability of the final model. Three broad classes of environmental variables were tested. The first group were unmodified climate variables such as mean annual precipitation and minimum monthly temperature. The second were environmental indices which were redefined by known physical processes to more closely reflect gradients of resource supply that directly influence plant performance. A soil water balance model was used as an example of this class. The third were physiological productivity indices and canopy carbon-uptake indices that were determined from genetic parameters of plant response to variation in environmental conditions of light and temperature.

It was found that where high quality soils information was available, the use of a soil water balance model provided a better estimate of species' performance and presence or absence than did the use of monthly rainfall and evaporation estimates alone. This water balance model used information about the water retention characteristics of the soil environment to distinguish site differences by the potential water supply. This estimation method was evaluated by comparison with physiological approaches to water balance modelling based on species' genotype responses and measurements of leaf area index. It was found to give comparable results and represents a simplification of the specification of water balance for the purpose of species' distribution modelling. However, in many cases the necessary soils information may not be available. Nevertheless, it was found that climatic based estimates of water balance that assumed all sites to have similar soil depths, and for which estimates of texture differences could be approximated from parent rock types were useful for some applications. This was highlighted using a case study in which the simple soil water balance model was used to explore the coexistence of two species of eucalypts, *E. obliqua* and *E. tenuiramis*. A water supply gradient allowed the comparison of data collected at different scales of study. However, for the purpose of deriving realised niche models, the use of these water-balance indices did not improve predictive performance, indicating that regression analyses of species' distribution will be limited by the quality of soil information typically available with compiled ecological data.

An average of the known physiological performances of eucalypts was used to develop productivity indices related to canopy carbon uptake from the environmental data available in the compiled dataset. These indices were generally poor predictors of species' occurrence. It was concluded that light and temperature, which these indices were designed to supplement, were themselves direct environmental factors and that little information gain was likely. In addition, their combination into productivity indices only introduced the potential for inappropriate assumptions that acted to mask species' differences. However, it was found that these productivity gradients could be used to address comparative questions related to individual species' realised and fundamental niche responses, and possibly the niche relationships between species. This was performed by considering the position of their ecological optima defined from environmental indices but arranged along these physiological gradients in productivity.

The methodological aspects of developing realised niche models were further considered for the distribution of *E. globulus* in eastern regions of Tasmania. An hierarchical approach to analysis was taken, considering the extensive literature for the ecology, genetics, silviculture and physiology of this species. The ecological analysis considered pattens of co-occurrence between *E. globulus* and other species within its geographic and environmental range. Initially univariate responses of the dominance and occurrence were derived followed by multivariate regression analyses of up to 39 candidate biotic and abiotic environmental gradients, and with up to their fourth order polynomial functions. With this number of starting variables, forward selection methods were unrewarding and a backward elimination procedure was quickly adopted. It soon became apparent that the closely constrained sampling domain did not allow for adequate definition of the absence response. The sampling domain was therefore extended to include all occurrences within the envelope of the geographic range defined by the atlas of distributions plus a 10 km range extension without any altitude restrictions. Model performance improved dramatically, as did the fits to the univariate responses.

It was concluded that the need to constrain a sample to avoid the problem of 'naughty-noughts' was not warranted in multivariate models that explain a substantial proportion of the species' response. Rather, prediction errors in interpolation or extrapolation of a species' response, apart from the problems inherent to the use of polynomials, were considered to be a problem of sampling bias and/or the exclusion/inclusion of inappropriate predictors.

Subsequent direct gradient analyses of the realised niche models revealed that the ecological optima and range of environmental gradient responses could be related to key physiological processes observed in field experiments of plantation grown *E. globulus*. The models also explored the possibility that the ecological response of *E. globulus* in wet or dry forest stands might be indicative of two genetically divergent ecotypes identified by racial classifications.

To explore the question of competition with a sympatric species, different models were defined for the response of *E. globulus* or species of the white gum complex (*E. viminalis*, *E. dalrympleana*, *E. rubida*) based on their respective pure stand or mixed stand occurrences. The ecological ranges or optima derived by direct gradient analyses of the realised niche models indicated that the mixed species stands did not necessarily occupy intermediate environments when considered as responses arranged along one or two gradients. Rather, mixed stands dominated by either species appeared to occupy distinct types of environments from each other and from those in which the respective pure stands occurrences were found, although these were ecologically closer to the respective pure stand response.

These patterns indicated that complex ecological relationships existed between species because of a probable hierarchy of interactions between a large number of potentially coexisting species and their patterns of response to each other and environment. Possibilities for considering the structures of these hierarchies could be explored using approaches similar to the predictive modelling and gradient displays developed in this thesis and demonstrated for some aspects of the ecology of *E. globulus*.

Acknowledgments

The work presented in this thesis initially arose as a continuation of a project for predicting rare vascular plant species' distributions in Tasmania's production forests. This project was instigated by Mick Brown, leading to the project work by Fred Duncan in 1988 (see Duncan 1988), and myself in 1990 as an honours student in the Department of Plant Science, University of Tasmania (see Williams 1990).

This thesis was jointly supervised by Prof. J. B. Reid (Department of Plant Science, University of Tasmania), Dr. M. J. Brown (Forestry Tasmania) and Dr. M. P. Austin (CSIRO Wildlife and Ecology, Canberra). Much of the data used in the development of predictive models in this thesis were made available for analysis by Forestry Tasmania (see Table 2.1). This work was supported by an Australian Commonwealth Department of Primary Industries & Energy Forestry Postgraduate Scholarship and equipment grant.

Two publications arising from work undertaken toward this thesis are attached, and acknowledgments of persons and institutions involved in those projects were given there (Battaglia & Williams 1996; Williams & Potts 1996). The first publication is also presented in Chapter 5 and arose from a collaboration with Mike Battaglia during his ARC research in the similar field of eucalypt ecology and species' coexistences. The second publication refers to the eucalypt atlas which was compiled with the close collaboration of Brad Potts.

The final editing and production of this thesis was made possible by the efforts and commitment of Dave Heatley, Sue Rundle, Jon Marsden-Smedley, Stephen Mattingley and Jayne Balmer. Their enduring support and encouragement is greatly appreciated.

Many thanks also to Mike Battaglia, Mick Brown, Brad Potts, Jim Reid and Tim Rudman who assisted with their comments on manuscripts at various stages of the final drafting of this thesis.

Statistical advice was sought from a broad range of people as the need arose. Initial approaches to the application of generalised linear modelling to predict species' distributions within the context of ecological theory were guided by Mike Austin. I am very grateful for his time in teaching me about these statistical applications and for the many publications that document his work that were made available while some were in early stages of drafting and development. Mike Austin also suggested using WATBAL for modelling site water balance.

The undergraduate courses in applied statistics and the concurrent training in the use of SAS by Glen McPherson were important prerequisites for undertaking these analyses. Thanks to Glen for his continuing encouragement, advice and interest in this project. More recently, I would like to express my gratitude to David Ratkowsky for many interesting and challenging discussions on different aspects of statistics, especially logistic regression, generalised linear and non-linear modelling, and some obscure methods that I managed to dig out of the literature and apply with his assistance.

Thanks also to Mike Battaglia, Brad Potts, Greg Jordan, Nuno Borralho, Wendy Catchpole and Steve Candy who were also approached at various stages of development of ideas for summarising, analysing, modelling or presenting the results of predictions based on ecological data. In particular, thanks to Mike Battaglia who always made time to listen to my ideas and offer alternative suggestions or patiently explained why it wouldn't work or helped out with mathematical approaches and SAS programming to build the idea into something bigger. Thanks again to Mike Battaglia for bequeathing his slightly faster 486 with much more disk space, without which this work could not have been completed. Subsequently, I am indebted to Brad Potts and Greg Dutkowski who made available an extra gigabyte on a partitioned NT drive, and saved me from much time-consuming desktop management. However, the final niche models could not have been run without the support of Greg Dutkowski who made his very, very much faster computer available to me while on leave for the month of September 1997.

This work has also touched many others. In particular I would like to thank Andrew McDonald for his friendly computing support, and the University of Tasmania IT staff who were always prepared to deal with problems as promptly as possible. Many thanks too, to the anonymous SAS technical support staff, just a phone call or an email message away.

The ecological data and climate estimates used in this thesis were originally supplied through staff at Forestry Tasmania. Thanks to Simon Orr, Bruce Haywood, Kevin Hortle and Sonja James for their assistance in the provision of these data. More recent estimates for raindays and corrected estimates for solar radiation were kindly provided by Darryl Mummery (CSIRO Forestry, Hobart). In particular, I am indebted to Darryl Mummery for promptly summarising and supplying the Tasmanian-wide estimates for climate in 1 km^2 grid cells, without which the direct gradient analysis displays and predicted distributions would not have been worth presenting. Thanks also to Greg Dutkowski and Mike Battaglia for facilitating the transfer of this data and making it available to me for use in this thesis.

Finally, I would like to thank my many friends and colleagues, too many to name individually, who have watched me grow with this thesis, offering much encouragement and support and many lasting happy moments. But especially I would like to thank Jon Marsden-Smedley who suffered the good bits and the bad, but has unconditionally supported me all the same.

Table Of Contents

Figures

Tables

Boxes

Maps

1. Predicting plant distribution patterns: a literature review

This chapter reviews links between recent physiological research and ecological theory as a guide to the development of a systematic approach to the correlative analysis of plant distributions.

1.1 Introduction

Land managers often exploit site variability as a means of optimising productivity. They may use models of plant-environment relationships to match species to the sites on which they grow best, or to predict the effect of changing a particular land management regime. Detailed experimental work in plant ecology and ecophysiology has provided the theoretical basis for developing robust and precise models for growing individual species under specific conditions or estimating probabilities of occurrence. Unfortunately, it is logistically not possible to study intensively all species in a timely and cost-effective manner. An interim strategy for evaluating plant performance is therefore needed (e.g. see Norton & Williams 1992; Ehrlich 1996).

With inventories of species' distributions and associated environmental data, predictive models of plant performance can be developed rapidly using correlative statistical techniques. To be useful, these models need to be based on sound ecological theory, and be consistent with the statistical assumptions of the analysis (Austin *et al.* 1990; Austin & Meyers 1996). Predictive modelling therefore provides the link between inventory data and questions of land use planning and management (e.g. Norton & Williams 1992; Prance 1994; Stork 1994), as well as opportunities to explore hypotheses about the association of plant species and ecosystem processes (e.g. Austin & Smith 1989; Austin 1991b; Neilson *et al.* 1992; Austin & Gaywood 1994). The predictions can subsequently be used to guide and focus detailed observational studies of population dynamics and/or experimental designs (e.g. Carey *et al.* 1995b).

Some recent developments in ecological theory have implications for the way in which a correlative analysis of species' distributions may be undertaken (e.g. Austin & Smith 1989). These developments have arisen largely from physiological research on individual species and foreshadow new directions for ecological study (e.g. Chapin *et al.* 1996b; Steffen *et al.* 1996b). Previous ecological theory arose from an holistic study of patterns of species' occurrence within plant communities (e.g. Clements 1936; Watt 1947). A more recent reductionist approach, based on an understanding of physiological and population processes, is increasingly being applied to the prediction of plant distribution patterns (e.g. Smith & Huston 1989; Bugmann 1996b).

Correlative analyses of species' distributions provide an interim strategy for understanding plant-environment relationships, but are generally limited to interpolative predictions. However, if the

reductionist approach is used to guide the choice of response and explanatory variables in the model, then there is a greater potential for extrapolating the predictions.

1.2 The biological foundation of vegetation-environment correlation

1.2.1 Vegetation patterns

Vegetation patterns are highly correlated with environment, but different processes dominate at different scales. For example, Coughenour and Elis (1993) found that ecosystem structure in dry tropical environments was hierarchically constrained by physical factors: (i) by climate at regional to continental scales; (ii) by topographic effects on rainfall and landscape water redistribution, and geomorphic effects on soil and plant available water at the landscape to regional scales; and (iii) by water redistribution and disturbance at local and patch scales. Neilson *et al.* (1992) suggested that such vegetation-environment relationships are more than simply correlations, but are mechanistically founded in the water balance and thermal regimes of a region.

At continental and global scales, climatic extremes of temperature and moisture are important determinants of the distribution of major physiognomic types (Woodward 1987). Each physiognomic type is characterised by a different physiological response, according to its mechanisms of tolerance and sensitivity to chilling, freezing or desiccation (e.g. Sakai & Weiser 1973). Therefore, each physiognomic type has a competitive edge for a specific combination of climatic conditions (Woodward 1987; Woodward & Williams 1987).

At regional and local scales, the obvious correlation between assemblages of species and the environment led to the belief that plant communities are the fundamental unit of organisation in vegetation (e.g. Clements 1936). However, dynamic and transient associations of individual species also contribute to plant community diversity (Gleason 1939, Whittaker 1975). The individual plant response to the environment determines whether it may occupy a particular micro-habitat (e.g. Menges & Kimmich 1996; Rusch & Fernandez-Palacios 1995). Competition with neighbouring plants of different species will also influence the presence of individual plants, and ultimately, therefore, the population density of species (e.g. Burton & Bazzaz 1995; Hara *et al.* 1995). Changes in climate or the micro-habitat will lead to further changes in the way species interact, altering the structure and/or composition of the community (e.g. Aguilera & Lauenroth 1995; Bazzaz *et al.* 1995). The plant community is therefore a phenomenon of a particular time and location, at a particular scale of observation. It is therefore a convenient unit for summarising and communicating complex interactions (Gleason 1939; Austin & Smith 1989).

1.2.2 Species' distributions

Species' distribution patterns reflect the physiological response of individual plants to the environment. Prediction of distribution requires an understanding of these responses. However,

several generalisations are necessary to successfully predict species' occurrence and apply it to future scenarios of environmental type (Chapin *et al.* 1993b).

Firstly, each species shows a unique response to climate. Therefore, climatic change can cause species to migrate and/or form new associations (e.g. Nowak *et al.* 1994; Starfield & Chapin 1996; Watts *et al.* 1996). However, the individualistic response of species and their interactions with other species make it difficult to decide which aspects of climate are critical determinants of distribution. This makes prediction of of future distribution patterns difficult, especially when relationships are based on the correlation between species' occurrence and environment.

Secondly, changes in the distribution of a species may lag significantly behind climatic changes where there are limits to the rate of species migration (e.g. Prentice *et al.* 1991). Thus, prediction of the future response to climate requires some understanding of factors governing the regeneration phase (e.g. Battaglia 1997; Russell-Smith 1996).

Thirdly, plant response depends on interactions with other species in the community and on environmental factors other than climate. For example, individual plants are sensitive to the availability of soil resources and the combination of competitive or facilitative interactions with other organisms in acquiring limiting nutrients (e.g. Chapin *et al.* 1986; Turkington *et al.* 1993; Turkington 1996). Plant distribution is also related to barriers to dispersal and the distribution of predators and pathogens (e.g. Loehle & le Blanc 1996), as well as disturbance regimes, such as fire (e.g. Barton 1993). Therefore, knowledge of climate alone is not enough to project species' responses into the future or onto different sites. Furthermore, complex and diverse communities, such as dry or wet tropical forests, have so many key stone species that it is unlikely climate could be used as the primary mechanism for predicting future vegetation change (Chapin *et al.* 1993b).

To overcome the inherent difficulties in predicting species' distributions, many workers have turned to functional groupings of species as a means of reducing the inherent complexity that needs to be considered when defining scenarios of future response (e.g. Smith *et al.* 1993; Steffen 1996; Woodward & Cramer 1996). For example, functional type classifications have been used to predict vegetation patterns by simulating responses to the environment (e.g. Chapin *et al.* 1996a; Steffen *et al.* 1996a), and for interpreting a functional basis for species' responses from the correlation of their distribution patterns with environment (e.g. Rutherford *et al.* 1995). While simulation studies provide a means of testing the practical and theoretical application of a functional type classification, the lack of experimental data for individual species, but extensive floristic inventories, makes the interpretation of species' responses from correlative analyses an important interim analysis.

1.2.3 Operational environment

The prediction of plant species' distributions requires an operational definition of environment (e.g. Mason & Langenheim 1957; Waring & Major 1964; Emmingham 1978; Mooney & Chapin 1994). The environment of a plant is a reflection of a complex history of interactions between climate, substrate, disturbance, vegetation, fauna and other life forms (e.g. Major 1951). For practical reasons the description of environmental variation often requires the use of scalars, indices and surrogates (e.g. Loucks 1962; Austin *et al.* 1984, 1990; Nemani *et al.* 1993; Sullivan & Chesson 1993; Ferrier & Watson 1996; Faith & Walker 1996). The operational definition of environment and choice of explanatory variables associated with a response therefore depends upon the scale and purpose of the study.

For the purpose of modelling plant distributions, the choice of sample size depends upon the environmental space that is utilised by an individual of the target species and the neighbouring plants that may directly or indirectly influence it (e.g. see discussion by Kenkel *et al.* 1989; Tothmeresz 1995). For example, the appropriate scale for sampling the community and environmental context for a population of mature forest trees might be in the range 0.1 to 0.3 ha. This reflects the scale of experience of the environment by generations of interacting tree species at a site. In the case of floristic inventories, the performance of a population at a site may be simply interpreted as either the presence or absence of the target species. Many such observations at different sites represent a sample of the species' geographic distribution.

Samples of a species' distribution based on spatial observations of occurrence or relative performance are relatively simple to collect and large data sets can be efficiently accumulated from a range of sources (Austin 1991b). Despite their obvious value, fewer samples of within-site monitoring of the cyclical changes of species' response with season or disturbance exist because of the increased expense involved in establishing permanent plots, and the extra time required for data collection before analysis (e.g. Fahrig *et al.* 1994; Herben 1996; Debussche *et al.* 1996; Condit *et al.* 1995, 1996a). Therefore, descriptions of the environment are usually limited to spatial properties. However, temporal heterogeneity in environment caused by disturbances, such as fire, flood and landslide are known to influence the developmental stage of the vegetation (e.g. Barton 1993; Iwasa & Kubo 1995; Bornette & Amoros 1996). As a result, correlations of species' distributions with their current environment will have residual unexplained variation.

The biotic habitat associated with a species may be viewed as an emergent property of the vegetation (*sensu* Austin & Smith 1989; Austin & Gaywood 1994) reflecting the combined effects of spatial and temporal heterogeneity in the environment (Stevenson 1997). Factors of plant community structure and composition reflect these accumulated differences in environment between sites. For predictive purposes, biological attributes may be suitable surrogates for unexplained variation in habitat heterogeneity due to temporal effects, when used in combination with indices for spatial variability in the physical environment (e.g. climate and substrate

characteristics). In addition, because of the inherent variability in the position of a site within its cycle of disturbance and succession (e.g. trajectory models by Noble & Slatyer 1980, 1981; Moore & Noble 1993; Noble & Gitay 1996), replicate samples of similar types of habitat may be needed when using biological surrogates as explanatory variables in models of plant response.

The interpretation of successional patterns from plant distribution patterns may be possible across habitats that appear similar except for variation in the developmental stage of the vegetation. This corollary of relating species' distributions observed across many places at one point in time to indicate trends in successional processes and responses at one place has been a long-held premise of plant ecologists (e.g. Gleason 1939; Watt 1947; Whittaker 1967). This is much the same as using the spatially-correlated patterns of species' response to predict scenarios of changing vegetation and species' distributions with shifts in climate, disturbance or land use (e.g. Huntley *et al.* 1995). However, other studies suggest that the interpolation of temporal processes from spatial observations may not be appropriate, because the properties of species that determine their long-term spatial dynamics are not the same as those which determine behaviour during succession (e.g. Prentice *et al.* 1991; de Swart *et al.* 1994; Sykes & Prentice 1995, 1996).

In undertaking correlative analyses, the use of attributes for vegetation structure and composition as surrogates for the influence of temporal processes on the outcome of plant responses does not necessarily resolve the question of which attributes of climate, substrate or disturbance have been omitted in a spatial study of plant distribution patterns. However, deduction from a knowledge of which factors are important to plant distributions and which have already been included in an analysis may help resolve these issues. The following sections therefore review the physiological basis of plant response in more detail.

1.3 Physiological responses

The accumulation of physiological responses by many genotypes to their environment gives rise to species' distribution patterns. The physiological response curve of a plant reflects the differing environmental conditions in which survival, growth and reproduction are possible (Fig. 1.1). The form of the plant response to increasing levels of a factor (termed environmental gradient) is characterised by an optimum region, either side of which performance (e.g. abundance, growth, rate of photosynthesis) declines (e.g. Tenhunen & Westrin 1979; Wilson & Keddy 1985; Austin & Smith 1989; Austin 1992; Sultan & Bazzaz 1993a-c; Sands 1996).

The optimum response of a species to increasing levels of an environmental factor reflects the range of conditions in which plants grow and reproduce best and leave the most successful descendants (e.g. Jaindl *et al.* 1995; Barton & Gleeson 1996; Hansen *et al.* 1996). However, plants may also survive over a wider range of conditions than those in which they can grow (e.g. Dias-Filho & Dawson 1995) or may grow over a wider range of conditions than those in which they can reproduce (e.g. Burton & Bazzaz 1995). This resilience of response to varying

environmental conditions represents a functional homeostasis, involving the whole plant life-cycle and its population dynamics, aimed at maintaining reproductive fitness (Sultan & Bazzaz 1993c).

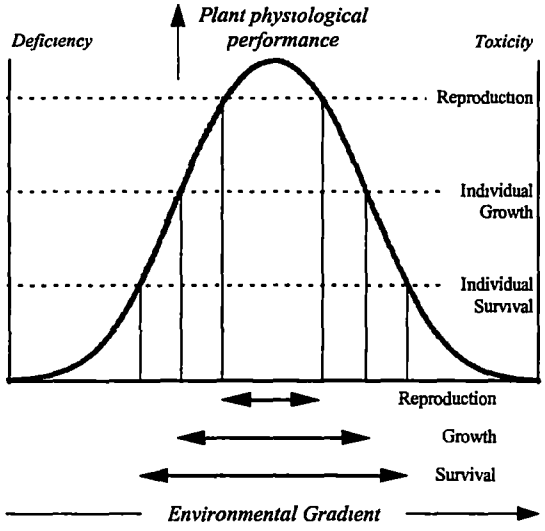


Figure 1.1 Generalised system of plant response to an environmental gradient (redrawn from Begon *et al.* 1990).

Response shape may be skewed, broadened, narrowed or otherwise varied from symmetric, but is always characterised by an optimum region, either side of which there is a declining response and limits to long-term survival, including extreme conditions that can be tolerated for short periods.

The range of ambient conditions in which reproduction occurs therefore defines the conditions for continuous existence of a population: the unit of observation in species' distributions. These conditions need only occur for as long as it takes to complete a stage in the reproductive cycle. For example, annual plants need only a relatively brief period of optimal conditions within a growing season to complete their whole life cycle (e.g. Tielborger & Kadmon 1995), but perennial species need only allocate sufficient resources to complete flowering and seed set (e.g. Gunster 1993; Stanton *et al.* 1997). The success of either response may in part be due to the timing of the presence of pollinators or dispersers, or to the suitability of conditions in the preceeding season or seasons as in the case of masting (e.g. Haase *et al.* 1995; Herrera *et al.* 1994; Waller 1993; Sork 1993; Appanah 1993). In addition, individuals of a species may persist in marginal habitats where they can survive but not reproduce, providing there is a compensatory inflow of propagules (Shmida & Wilson 1985), or they may reproduce less prolifically but are nonetheless genetically and ecologically important to the species as a whole (Levin 1995).

1.3.1 Relationship between species' distributions and population performance

The size of species' geographic ranges and their average local abundances at sites where they occur are positively correlated (Holt *et al.* 1997). This phenomenon of species' distributions has been termed the range-abundance correlation (e.g. Gaston 1996a, b; Gaston *et al.* 1996). In the case of plant species, their ecological performance, which may be recorded as the presence or absence of individuals within a population for a range of populations across many habitats, could be viewed as an indicator of the relative performance of individuals within a single population, and *vice versa* (e.g. Paruelo & Lauenroth 1996; Naeem 1996). The range-abundance correlation may be due to habitat availability: the amount of suitable habitat within a site is likely to be higher within core areas of a species range and lower in marginal areas of a species range (e.g.

Venier & Fahrig 1996). However, in the case of some widespread, locally rare species (e.g. see classification of Rabinowitz 1981), a simple range-abundance relationship may not apply because indeterminate factors of migratory history and environmental change may have little relationship to current habitat conditions (e.g. *Eucalyptus cordata*, Potts 1988; and other spatial studies of species' rarity by Rabinowitz *et al.* 1986; Schwartz 1993; Saetersdal 1994).

The species range-abundance relationship has important implications for the interpretation of estimated probabilities of occurrence from species' distribution patterns. Different physiological and ecological mechanisms may contribute to different patterns of range-abundance correlation (Holt *et al.* 1997). Therefore, by analogy, the characteristic response curve for a species based on its distribution could be interpreted as reflecting the accumulation of conditions leading to a given level of reproductive success, growth and survival by individuals within any one population. That is, relative measures of species abundance, fitness, density or population viability could be interpreted from modelled relationships. This correlation has implications for conservation evaluation, such as predictions for the occurrence of rare species (Ferrier & Pearce 1996), assessment of extinction risk (e.g. Burgman *et al.* 1993) or reserve selection and design (e.g. Pressey *et al.* 1996). Performance estimates related to growth rather than survival may also have application to site selection for agriculture and forestry purposes (see contrasting examples of site selection for forestry by Lindenmayer *et al.* 1996 and Battaglia & Sands 1997).

1.3.2 The concept of environmental gradients

Environmental gradients are simply the variables which may be associated with plant performance, and which can be arranged on a continuous or ordinal scale from low to high levels. At least two categories of environmental gradient can be directly related to a plant's physiological response (Austin 1980). These are the resources consumed by plants when growing, and conditions that govern growth rates and the maintenance of plant physiological integrity (Austin & Smith 1989; Begon *et al.* 1990). Indirect environmental gradients, such as altitude and latitude, are not suitable for a direct analysis and interpretation of plant responses from species' distribution patterns (Austin & Smith 1989).

Plants require relatively few resources for growth. Apart from space, these are light, water, carbon, oxygen and mineral nutrients. Ambient conditions, such as temperature and solute redox potential (e.g. measured as the concentration of hydrogen ions, expressed as pH), influence rates of resource acquisition and metabolism in complex ways (e.g. Delucia *et al.* 1997; Criddle *et al.* 1996a; Ingestad & Agren 1996).

The close interaction between environmental gradients and their coupling in plant physiological processes results in some factors that are resources at one level acting as conditions at another level. Some resources act as conditions by indirectly influencing plant function through their effect on the supply of other resources. For example, drought conditions limit water acquisition but also reduce the supply of mineral nutrients, especially nitrogen (e.g. Hawkins & McDonald

1994). Drought may also limit the acquisition of carbon for photosynthesis due to stomatal closure, and it may reduce leaf size by lowering the turgor pressure available for cell expansion (e.g. Raison & Meyers 1992). Waterlogging represents excess or toxic conditions along a water gradient (e.g. Wang & Klinka 1996). It limits the supply of oxygen to roots and, over extended periods, may reduce root function and therefore the acquisition of water and nutrients. This is similar to imposing a drought effect.

Soil pH determines the chemical state of mineral nutrients and therefore influences the availability of minerals to the plant. In alkaline soils, iron, manganese, phosphorus and certain trace elements are fixed in relatively insoluble compounds, limiting their availability for uptake by plants and resulting in deficiencies (Binkley & Vitousek 1989). In acid soils, the increased availability of some ions, such as aluminium, can be toxic to higher plants (Begon *et al.* 1990). Therefore nutrients may also be available to plants as resources which promote growth or as conditions that induce toxicity.

Plant distributions and growth rates are limited by seasonal and diurnal high and low temperature extremes and also by the average temperatures experienced (Criddle *et al.* 1994, 1996b). Very high or very low temperature conditions influence cellular integrity and the water status of the plant (e.g. Larcher 1995). For example, high temperature conditions create a demand for evaporation that may desiccate plant tissues (e.g. Kitano & Eguchi 1993), or deactivate enzyme systems. Low temperatures also influence rates of cellular metabolism, reducing the uptake of resources and their transfer across cell membranes (e.g. Hallgren & Oquist 1990). Chilling and freezing temperatures also create desiccating conditions and may be lethal when ice particles damage cell structures.

The diurnal variation in temperature means that the extremes are typically associated with different physiological processes. For example, overnight temperatures may have a greater influence on metabolic processes, such as the rate of dark respiration (e.g. Hansen *et al.* 1994; Criddle *et al.* 1994). On average, the effect of winter temperatures may be dominated by the degree of overnight low temperature stress, and the summer temperatures may largely influence day- and night-time levels of respiration and subsequent carbon gain (e.g. Criddle *et al.* 1996b; Hansen *et al.* 1996).

Sparse vegetation cover is frequently associated with stressful environments, and therefore plants must also cope with the inhibitory effect of high light intensity, suppressing photosynthetic activity (e.g. Pugnaire *et al.* 1996). For example, photoinhibition due to high light intensity is exacerbated by thermal extremes when the photosynthetic apparatus is active during the daytime (e.g. Ogren & Evans 1992; Eilers & Peeters 1993). In general, when plants are already stressed by high or low extremes of water, temperature or nutrient supplies, the added stress of high light intensity may have a damaging or inhibitory effect on the photosynthetic apparatus. The adverse

effect of high light intensities depends on the plant's ability to acclimatise to changing conditions (e.g. Claussen 1996).

It is therefore difficult to always separate the function of environmental gradients as resources that are integrated into plant parts from the conditions that facilitate or inhibit plant metabolic processes. A different perspective is needed.

An environmental gradient that has a direct influence upon whole plant growth or metabolic processes may be considered as comprising two functions with respect to the physiological response (Fig. 1.2). The first is the plant-productivity function. It includes conditions that affect the rate of uptake of plant resources and is represented in the optimum region of the plant physiological response curve. Levels of an environmental gradient which have a productivity function are likely to lead to plant growth (all else being optimal) and successful reproduction, leading to the long-term perpetuation of a species at a site. The second function, the plant-stress function, is directly or indirectly inhibitory to the accumulation of plant biomass. This may be due to disturbance factors, such as fire; or to predation, including herbivores, pathogens and parasites. Physiological stress is indicated by the decrease in plant performance below its optimum (e.g. Osmond *et al.* 1987; Lichtenthaler 1996).

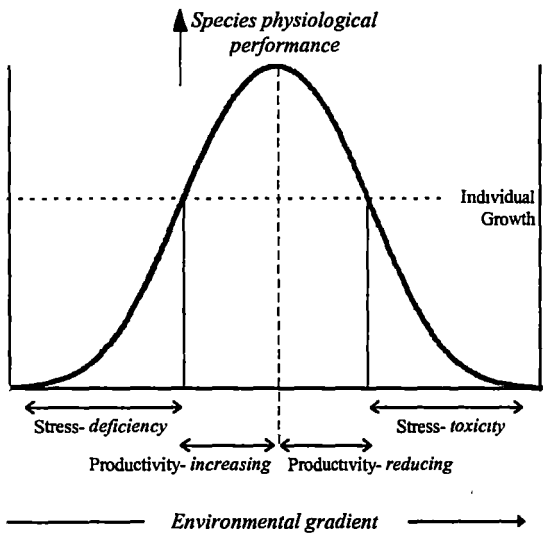


Figure 1.2 Plant-stress and plant-productivity functions of an environmental gradient defined relative to the generalised physiological response of a species.

At different positions along an environmental gradient, the physiological processes of plant response are likely to be different. The two-sided characteristic of the physiological response means that each environmental gradient has two productivity effects and two stress-effects.

Since the characteristic plant response is two-sided, the productivity and stress functions of an environmental gradient will each comprise two regions associated with this response (Fig. 1.2). The stress function of an environmental gradient is the main determinant of species' distributions (e.g. Osmond *et al.* 1987). Growth functions subsequently determine core habitat and the distribution of relative performance (e.g. Criddle *et al.* 1996b).

Although the division between a plant-stress function and plant-productivity function of an environmental gradient appears somewhat arbitrary, the physiological processes involved in whole-plant response can be expected to differ in importance at different positions along the gradient. For example in cool temperate habitats, the plant response to a temperature gradient

may be largely due to tolerance of cellular freezing during winter periods, or to the ability to maximise carbon acquisition during the summer periods.

In addition, environmental gradients are themselves codependent. Factors, such as water and temperature, interact and co-vary in a number of different ways. On a landscape scale, interaction between climatic gradients in temperature, water and light vary with region and terrain, indices for which may be interpolated from seasonal measurements among representative weather stations (e.g. Hulme *et al.* 1995, 1996; Hutchinson 1987,1989; Lehman 1987; Law & Waring 1996).

The correlation between plant distributions and environmental gradients needs to account not only for the characteristic form of the physiological response, but also for the interdependence between measured environmental factors. A large number of indices may be needed to adequately define these relationships. The role of stress due to particular levels of a range of environmental gradients in determining the probability of occurrence of a species is therefore explored in the following sections.

1.4 Response to stress

Most plants respond to stress in similar ways and require similar resources, but adaptation to different intensities and frequencies of stress has led to different life-history strategies and phenotypic variability. These mechanisms are a means of mitigating stress (Sultan 1987).

The progression from benign to severe or stressful environments or from favourable to unfavourable seasons in a single locality is correlated with a reduction in potential plant productivity (Mooney 1980). This is because selection favours plant growth rates close to the maximum that can be maintained in any given environment (Chapin *et al.* 1987, Chapin 1991). In low-resource environments, a low potential growth rate is an adaptive response linked to the capacity to accumulate reserves during rare periods of high-resource supply for later support of growth during periods of low-resource supply (Bazzaz *et al.* 1987; Chapin *et al.* 1990, 1993a). However, in resource-rich environments, plants tend to have a high potential growth rate and lower resource-use efficiency.

The integration of the plant response is largely mediated through the physiological couplings of water, light, carbon and nitrogen in photosynthesis and the respiratory costs of their acquisition and allocation to biomass (e.g. Osmond 1987; Schulze *et al.* 1987; Chapin *et al.* 1987; Dewar 1996a, b; Kaitaniemi & Honkanen 1996). The close relationship between carbon and nitrogen, in particular, can lead to dramatic changes in productivity with small changes in nitrogen availability (Tateno and Chapin 1997).

These source-sink mechanisms allow plants to adjust to an internal resource imbalance by temporarily allocating new biomass to the organs that acquire the largest amount of the limiting

resource. Alternatively plants may reduce the rate of tissue loss, due to factors such as herbivory or climatic extremes, aiming to restore an internal balance of nutrients that is favourable for growth (e.g. Bazzaz *et al.* 1987; Chapin *et al.* 1990). However, a fluctuating environment generally prevents plants from achieving perfect equilibrium with respect to resources (Chapin *et al.* 1987).

The rapidity and flexibility with which plants can respond to fluctuations in their environment lead to the proposition that plants have a hormonally-mediated, centralised stress-response system (Chapin 1991; Chapin *et al.* 1993a).

These hypotheses of integrated plant response systems have implications for interpreting the correlation between species' distributions and their environment. For example, the position of a plant species' optimum response at the dry end of a soil water supply gradient could be interpreted in terms of potential physiological and morphological traits required to maintain a viable population. If the same species has a distribution associated with low nutrients in addition to low water supply, then the stress response system suggests that specialist traits have evolved (e.g. see Table 1.1). Conversely, if a species occupies a habitat characterised by only one major gradient of stress (rather than multiple stress gradients), then the plasticity inherent to its genome may be sufficient such that specific adaptations of the stress response system are not required (Chapin *et al.* 1993a).

Table 1.1 Generalised suite of traits found in most low- and high-resource environments and specific adaptations to particular environments (reproduced from Chapin *et al.* 1993a).

	Habitat Type			
Traits	Infertile	Dry/saline	Shaded	High resource
Generalised adaptations:				
Maximum relative growth rate	Low	Low	Low	High
Photosynthetic rate	Low	Low	Low	High
Tissue nitrogen	Low	Low	Low	High
Stomatal conductance	Low	Low	Low	High
Transpiration rate	Low	Low	Low	High
Carbon-based defences	High	High	Low	Low
Nitrogen uptake potential	Low	Low	...	High
Root:shoot ratio	High	High	Low	Low
Leaf turnover rate	Low	Low	Low	High
Specific adaptations:				
Root phosphatase	High	...	...	Low
Osmotic adjustment	...	High	...	Low
Use of sun flecks	...	...	High	Low

1.4.1 Ecological significance of stress

Different types of adaptive response to variation in environmental conditions, as experienced by the populations of a species in different habitats, have implications for the interpretation of the response in terms of generalist or specialist physiological and morphological traits (e.g. Sultan 1995; Lortie & Aarssen 1996).

Phenotypic plasticity is an evolutionary response to environmental heterogeneity and is viewed as a major outcome of adaptation in populations (e.g. Sultan 1995; Via *et al.* 1995; Dejong 1995; Zhivotovsky *et al.* 1996). Phenotypic plasticity enables plants to maintain fitness by avoiding mortality due to environmental stress (including density stress) so that effects are absorbed within the generation (Sultan 1987). A species therefore reflects the sum of the phenotypes of individuals that comprise each population, and for modelling purposes, can correctly be treated as a super-genotype. Lortie and Aarssen (1996) suggest that the evolution of specialisation to the more favourable end of an environmental gradient can be expected to increase the plasticity displayed across habitats. Conversely, specialisation to the less favourable end of an environmental gradient can be expected to decrease the plasticity displayed. That is, there is a cost in phenotypic plasticity of specialisation to the less favourable (more stressful) types of environment.

Since individual populations of a species are locally adapted to their normal environment, this cost becomes apparent in species' distribution patterns. When considering the correlation between species' distributions and environment, the cost of specialisation to limiting levels of an environmental gradient may appear as a peaked optimum response with a narrow tolerance range. The lower levels of phenotypic variability are expected to equate with smaller ecological or environmental ranges.

Since a species may be viewed as a super-genotype, performance (e.g. relative growth rate, abundance, relative dominance, probability of occurrence) along an environmental gradient can also be viewed as a reaction norm reflecting reciprocal adaptive responses of the genotype and its constituent genetic variability (cf. Ellner & Hairston 1994). Therefore, the phenomenon of a cost associated with the degree of genetic specialisation in the extremes of low productivity environments (e.g. Lortie & Aarssen 1996) is analogous to the hypothesis of physiological trade-offs associated with a stress-response syndrome (Chapin *et al.* 1993a).

Different plant life-cycle stages may express different levels of stress from the same set of environmental conditions. For example, some desert plants maintain inter-annual seed banks as a temporal buffer against conditions that cue the germination response, but which may not lead to the successful completion of the life-cycle in a season (Pake & Venable 1996). In addition, different biomass allocation patterns between seedling and adult morphs of a species may exist as adaptive plastic responses to resource gradients, rather than fixed species attributes that determine growth rates (Gleeson & Tilman 1994). Furthermore, some species adopt long-distance clonal spreading as a mechanism which reduces the time needed to colonize recently and frequently disturbed sites, in which the environmental stress effect is probably too severe for reproduction and seed set, but is below the time needed for stem propagation (Fahrig *et al.* 1996).

Therefore, within each developmental phase, the plant has a specific set of ambient conditions which gives an optimum growth response (e.g. Claussen 1996; Kruger & Reich 1993). This

optimal response may vary continuously, as in the case of seasonal acclimation of photosynthetic temperature optima in frost-free habitats (e.g. Sall & Pettersen 1994; Spencer *et al.* 1994; Sands 1996). Alternatively, a threshold in photoperiod response may invoke, for example, dormancy or bud burst in deciduous species, or flowering and other developmental changes in non-deciduous plants (e.g. Salisbury 1981; Morgan & Smith 1981; Koski & Sievanen 1984).

At the global scale, the accumulation of specialisations to deal with stress contributes to the separation of life-forms and vegetation types on major climatic gradients for temperature and water (e.g. Holdridge 1967; Whittaker 1975; Box 1981a, b, 1996; Woodward & Williams 1987). Thus the generalisation of physiological attributes of plant response becomes a suitable basis for grouping or regrouping of species into functional types for addressing global and regional questions of vegetation distribution (e.g. Rutherford *et al.* 1995; Woodward *et al.* 1995; Haxeltine and Prentice 1996a; Woodward & Cramer 1996).

1.5 Concepts of fundamental and realised niches

The distinction between the fundamental and the realised niche is important when modelling plant responses. The observation that a species' performance in monoculture tended to be greater than in mixture led to the distinction of two types of response (Hutchinson 1957, 1961; Ellenberg 1953 in Mueller-Dombois & Ellenberg 1974; Austin & Austin 1980; Austin 1982; Fresco 1982; Parrish & Bazzaz 1982a, b; Austin *et al.* 1985). In the absence of negative interactions with other species (competition and predation), the complete array of physiological responses expressed by a plant to its multi-dimensional environment defines its fundamental niche. The corresponding ecological responses of a species, in the presence of all other biotic and abiotic conditions which may restrict its performance, is an expression of its realised niche.

Mechanistic or process models of plant physiological responses are often designed to predict the potential performance of a species in the absence of biotic interactions, such as competition (e.g. King 1996; Running & Gower 1991; Running & Caughlan 1988). However, these process models tend to over-predict species' distributions because negative biotic interactions and historical landscape processes frequently prevent plants establishing in habitats where they could potentially grow (Loehle & le Blanc 1996). Therefore, simulation models that explicitly include ecological processes of competition and predation with physiological response parameters, provide a basis for testing our understanding of the relative roles of fundamental and realised niches in determining species' distributions (e.g. Smith & Huston 1989; Bugmann 1996b; Korol *et al.* 1996a, b; Keane *et al.* 1996a, b; Malanson 1997).

1.5.1 Fundamental niche responses

Facilitative interactions between species, such as mycorrhizal fungi, rhizobium bacterial nodules and insect pollinators, may increase the availability of resources above the passive rates of environmental supply, or increase the rate of fecundity and the probability of regeneration

success (Bazzaz 1987). Such positive interactions are viewed here as components of a species' fundamental niche. A healthy symbiosis, although incurring a cost, is more likely to enable the plant to withstand environmental stress and competition for limiting resources, being the realised conditions of its environment.

Pollinators and dispersers may also be viewed as biological gradients of a plant's fundamental niche. They contribute to the potential for a species to occupy a site continuously (Bazzaz 1987). Insect or bird pollinators are essential to the reproductive cycle of some flowering plants (e.g. Paton & Ford 1983; Feinsinger *et al.* 1991; Feinsinger & Tiebut 1991). Dispersal of seed by wind or insect and vertebrate species is also essential to the long-term survival of some plants and ensures the potential for occupying new sites is continuously tested (e.g. Gunster 1994; Leishman *et al.* 1992).

The definition of positive or facilitatory interactions between species as a mechanism of the fundamental niche is unclear in some circumstances because plants that benefit from their neighbours may also compete with them (Hunter & Aarssen 1988). For example, positive interactions between some species may lead to competitive superiority of one plant growing in the presence of another (e.g. Bengtsson *et al.* 1993; Callaway 1995). In addition, facilitative mutualism may exist between understorey herbs or grasses and overstorey shrub species in semi-arid environments through reciprocal improvements in micro-climate, soil properties and nutrient availability creating 'islands of fertility' (e.g. Ludwig & Tongway 1995; Haase *et al.* 1996; Pugnaire *et al.* 1996; Moro *et al.* 1997). In wetland habitats, some species may only grow in the presence of aerenchymous plants that can transport oxygen into the rhizosphere of water-saturated soils (Callaway & King 1996). Some of these interactions could also be viewed as facilitative pathways of plant succession (Connell & Slatyer 1977; Noble & Slatyer 1981). Nevertheless, commensal and mutualistic behaviour between different species or plant life forms epitomises the significance of the habitat context, such as structural vegetation attributes, in the fundamental niche of a species, and the difficulty of separating the potential from the realised.

In other situations, the realised niche of a species may appear to be greater than the fundamental niche due to a continuous unidirectional migration of doomed organisms (Thomson *et al.* 1996). This occurs where a sink habitat contains an apparently stable crop of propagules from a source habitat, but where the organisms are unable to survive to reproduce (e.g. Pulliam & Danielson 1991; although see Fahrig *et al.* 1996).

1.5.2 Realised niche responses

Empirical prediction of a species' distribution from information about a species' natural occurrence and its associated environment represents a model of its realised niche (e.g. Austin *et al.* 1984, 1990, 1994; Bartlein *et al.* 1986; Yee & Mitchell 1991; de Swart *et al.* 1994; Stahle & Chaney 1994; Leathwick 1995; Huntley *et al.* 1995; Austin & Meyers 1996; Wu & Huffer 1997). It is simpler to describe the contemporary physical environment of a species, yet we know

that occurrences are due to the cumulative effects of competition, predation, disturbance and historical climates or other idiosyncratic conditions (e.g. Wardell-Johnson & Horwitz 1996).

In some cases, species' interactions may act to stabilise community species composition in the presence of moderate environmental change (Hattenschwiler & Komer 1995). In addition, since species perform differently at different sites, their interaction with other species will also vary between sites, further contributing to spatial variation in species' coexistence patterns observed as community species' composition (Kobe 1996). Thus a species' ecological response (an element of its realised niche) arises from constraints imposed on its physiological response (the corresponding element of its fundamental niche), and therefore ecological responses are inherently nested within physiological responses (e.g. see theoretical discussion by Austin & Smith 1989). This nested arrangement has implications for the interpretation of species distribution patterns (see Section 1.6, this chapter).

Changes in species' performance in response to different site conditions, and the subsequent interactions between species, are critical to understanding the processes of plant community resilience to climate change. Therefore, the role of biotic interactions (including competition, herbivory, pathogens and parasites) and historical landscape processes (including disturbance regimes) in the realised niche response of a species need to be considered.

1.5.2.1 Competition and predation

Interference interactions, such as competition and predation, reduce the level of resources available to plants, preventing their acquisition or removing assimilates as biomass (e.g. Schoener 1983; Connell 1983; Bazzaz *et al.* 1987). Only interspecific competition is considered here, because interbreeding dampens the effective outcome of the selection pressures arising from intraspecific interactions.

Competition and predation are believed to influence the environment of a plant by limiting the effective supply of resources (e.g. Goldberg 1990). This leads to a trade-off response by the plant that is fundamentally aimed at correcting an internal resource imbalance (Chapin 1991; Chapin *et al.* 1993a). For example, plants tend to preferentially develop those above or below-ground organs that are most likely to maintain access to limiting resources (e.g. Chapin *et al.* 1987; Burton & Bazzaz 1995). This response could also be aimed at protecting biomass from herbivores (e.g. Bazzaz *et al.* 1987; Chapin *et al.* 1990; Loehle 1996). Plants therefore need to be able to respond to the phenological cycles of foraging by other plants (e.g. Le Roux *et al.* 1995; Wilson & Tilman 1995), or the predatory cycles of herbivores, parasites and pathogens (e.g. Aizen & Patterson 1996).

Competitive interactions involve an effect and a response (Goldberg 1990). One species may create a competition-effect toward another species by suppressing resource levels, but that species may respond to competition through its ability to tolerate suppression or stress due to low resource levels (Bengtsson *et al.* 1994). Interactions between individual plants actually occur

through some intermediary, such as resources, pollinators, dispersers, herbivores, or microbial symbionts (Goldberg 1990). For example, Murphy & Aarssen (1996) demonstrated allelopathic effects of pollen from the grass, *Phleum pratense*, on pollen germination and seed set among other *Poaceae* species within the same local habitat.

Differences in population behaviour or spatial disposition among plants (e.g. dispersed versus clumped individuals, or exposed versus cryptic positions) may also result in one species having an indirect competitive advantage over another (e.g. Burger & Louda 1996). Similarly, the physiological cost of defence and the unpredictability of predator attack, may have lead to the evolution of inducible defence mechanisms in plants (Bazzaz *et al.* 1987).

Many different physiological mechanisms are involved in competition, depending upon whether a species is responsible for a competitive effect or whether it is the target of a competitive effect to which it responds (Aarssen 1989, 1992). Competition therefore takes many forms, the reciprocal effect of which may be symmetric or asymmetric, apparent or real (Connell 1990).

Coexistence between species may therefore be maintained in the presence of competition through a number of mechanisms (Aarssen 1992), such as shifts in competitive advantage with life-form stage (Bengtsson *et al.* 1994), balanced competitive ability (Aarssen 1983) or temporal cycles which maintain environmental heterogeneity through seasonality or disturbance regimes (e.g. see contributions in Kolasa & Pickett 1991).

Experimental evaluation of competition, to distinguish the mechanisms of effect and response, is necessary if we are to understand the determinants of individual plant fitness and community structure that lead to species' distribution patterns (Goldberg 1990, 1994, 1996; Goldberg & Barton 1992; Goldberg & Scheiner 1993; Goldberg *et al.* 1995).

1.5.2.2 Ecological consequences of biotic interactions

As previously noted, the physiological consequence of biotic interactions, whether obtrusive, facilitative or predatory, is one of multiple limiting resources (Chapin *et al.* 1987, 1993a). The source of a resource limitation therefore has consequences for the tactics and traits employed by the plant to restore an internal resource balance effected through flexible patterns of biomass allocation to growth, storage, reproduction and defence (Bazzaz *et al.* 1987; Chapin *et al.* 1990; DeAngelis & Huston 1993). These allocation patterns define a species' ecological role, and therefore plant distribution and adaptation (Bazzaz 1991). Furthermore, the influence of competition on foraging space and resource acquisition suggests that a relative (rather than an absolute) measure based on above- or below-ground interactions is more appropriate for comparing theories of community function (Grace 1993, 1995; Goldberg 1994; Reader *et al.* 1994).

For example, in high productivity habitats, especially mesic forests, above-ground competition for light influences individual plant architectures, such as aerial branch extension into adjacent canopy gaps (e.g. Coffin & Urban 1993; Tremmel & Bazzaz 1993, 1995; Bonser & Aarssen

1994; Walters & Reich 1996; Weishampel & Urban 1996). Similarly, below-ground competition for resources in low productivity habitats influences root architecture (e.g. Wilson 1993a, b; Coffin & Lauenroth 1996; Rodriguez *et al.* 1996). On a local scale, competition for water among neighbouring plants can be very intense (Schulze *et al.* 1987). Below-ground competition therefore becomes the dominant interaction in the arid or nutrient-poor habitats (e.g. Cromack 1981; Tilman 1988; Smith and Huston 1989; Huston & DeAngelis 1994; Belcher *et al.* 1995; Peterson & Squiers 1995; Briones *et al.* 1996; Lippert *et al.* 1996; Kremer & Running 1996). However, in habitats with chronic resource limitations, it is also likely that physical constraints (e.g. climate and soil nutrient status) will have a greater influence than biotic constraints on plant physiological responses (e.g. Twolanstrutt & Keddy 1996). Thus plants exhibit adaptive relationships between seedling root form and their soil environment (Neave & Florence 1994; Florence 1996).

These observations have been generalised to define a theory of competition for simulating plant distribution patterns. In general, the structure and composition of the vegetation is considered to be controlled by asymmetric (one-sided) competition for an above-ground resource (i.e. light), and symmetric competition for below-ground resources (i.e. moisture and nutrients) (e.g. Smith & Huston 1989; Huston 1992; Kropff & Spitters 1992; Nakashizuka & Kohyama 1995).

1.5.2.3 Niche differentiation

Theory predicts that intense competition between individuals with a high degree of overlap in resource use, over evolutionary time, may act as a selective force resulting in niche differentiation (Bazzaz 1990). A niche has therefore been broadly defined as 'the pattern of responses of an individual, a population, or a species to the physical and biological gradients of its environment' (Bazzaz 1987). However, within a species at a given location, a divergence in response is severely limited by gene flow. Niche differentiation through competition consequently tends to act between coexisting species, but complete divergence leading to the avoidance of competition may be rare (e.g. Aarssen 1989), depending upon the type of habitat (e.g. Bazzaz 1991).

Plant distributions reflect a myriad of different types of interactions between species and different stages of evolution toward specialisation or generalisation according to contemporary and historical patterns of competition, predation and abiotic stress. Overlapping distributions between species and different patterns of co-occurrence may therefore have implications for the ecological interpretation of niche differentiation from models of these responses (e.g. Bowers & Harris 1994). Methods of direct gradient analysis, in particular, provide a useful way of establishing the nature of the niche differences between coexisting species by considering their individual distributions as a continuum with respect to environmental gradients (Whittaker 1967).

1.5.3 Integrating experimental results and theoretical concepts

While correlative models of species' distributions represent approximations to the realised niche (Austin *et al.* 1984, 1990), in the absence of comparative information about the fundamental niche, it is difficult to interpret these ecological response patterns (Landsberg 1981). Explanation of species' distribution patterns therefore requires the integration of physiological and ecological studies with current theoretical ideas (e.g. Malanson *et al.* 1992).

Information about the relative competitive performance of two or more coexisting species, including the nature of any predatory interactions (herbivores, parasites and pathogens), could theoretically, enable their respective ecological responses to be predicted from their physiological responses (Austin & Smith 1989; Huston 1991). Comparative experimental studies of whole-plant species' responses in monoculture and mixture therefore provide a theoretical basis for the interpretation of plant distribution patterns, indicating the relative importance of competition or physiological limitations to growth, and the potential shapes of responses (e.g. Austin & Austin 1980; Fresco 1982; Austin *et al.* 1985; Grace 1988; Tremmel & Bazzaz 1993). The putative shapes of these responses, in particular, are important when selecting explanatory variables or testing functional relationships in ecological models (Austin 1987). Complex functions, including bimodal forms, may be necessary to describe some species' ecological responses to the constituent environmental gradients of its habitat (Fresco 1982; Austin *et al.* 1985; Bartlein *et al.* 1986; Leathwick 1995).

Experimental studies demonstrate the difficulty of measuring the proximal effect of one species upon the level of a resource available to another species. For example, reduced performance of *Cirsium vulgare* in mixture compared with monoculture can be interpreted as competition (Austin *et al.* 1985). Local nutrient depletion due to the presence of different species with different mechanisms of response might be the indirect cause of this competitive effect. The effective nutrient concentrations available to *C. vulgare* in mixture could be inferred by directly comparing its ecological performance with its physiological performance (e.g. the maximum performance of *C. vulgare* in mixture at 16 times normal nutrient concentration corresponds to a performance in monoculture at 4 times normal nutrient concentration — Austin *et al.* 1985). Differences in plant traits associated with strong effect and strong response competitive ability, especially in the early stages of competition, emphasise the importance of distinguishing between them in experimental studies (Goldberg & Landa 1991).

The depletion of nutrients from around plant roots increases soil spatial heterogeneity and hence opportunities for plant coexistence (e.g. Huston & DeAngelis 1994; Grace 1995). Such micro-habitat effects are also difficult to quantify. Therefore the observation of differences between species' physiological and ecological responses may, in part, reflect the mismatching of scales between the observation of plant performance and its environment (see Palmer & Dixon 1990). An understanding of this potential confounding effect in the analysis of species' distribution patterns enables the interpretation of responses to account for the fact that performance may not

be matched proximally to environment. Other factors, such as the degree of within-site variability, may need to be considered in empirical models to account for this difference in scales.

1.6 The continuum concept of plant niche

The continuum concept is a theory of plant response and the consequent effects for community structure relative to environmental gradients (e.g. McIntosh 1967, 1975; Austin 1985; Austin & Smith 1989; Collins *et al.* 1993; Austin & Gaywood 1994; Matthews 1996; Hoagland & Collins 1997). To compare and interpret species' ecological responses, Austin and Smith (1989) developed a series of propositions for the continuum concept that consider the shapes of species' fundamental niche relative to an environmental gradient. They distinguished physiological responses to temperature or pH as a series of independent, overlapping curves, and physiological responses to water, nutrients or light as a series of nested curves. Integrated physiological response models for individual species (e.g. Barton 1993; Running 1994; Nikolov *et al.* 1995; Friend 1995; Korol *et al.* 1996a; Walters & Reich 1996; Battaglia & Sands 1997) may provide the basis for testing these fundamental niche propositions (e.g. Malanson *et al.* 1992; Malanson 1997).

The physiological response of a species may take a variety of forms because different processes are involved at limiting or toxic levels of an environmental gradient (e.g. Austin 1992). This underlying shape has implications for the comparative interpretation of empirical models of plant distributions that approximate the realised niche (Fig. 1.3A). Without a knowledge of the underlying direction of skew, or the range of a physiological response, the relative effects of competition and other biogeographical processes cannot be inferred from an ecological response.

The ecological response parallels the physiological response and, with respect to the same environmental gradient and the same performance measure, it is equivalent to or nested within the physiological response (Fig. 1.3B). The ecological response varies in shape depending on its position along an environmental gradient, becoming increasingly skewed toward extremes of the gradient reflecting the relative importance of physiological tolerance and competition (Austin 1990; Austin & Gaywood 1994).

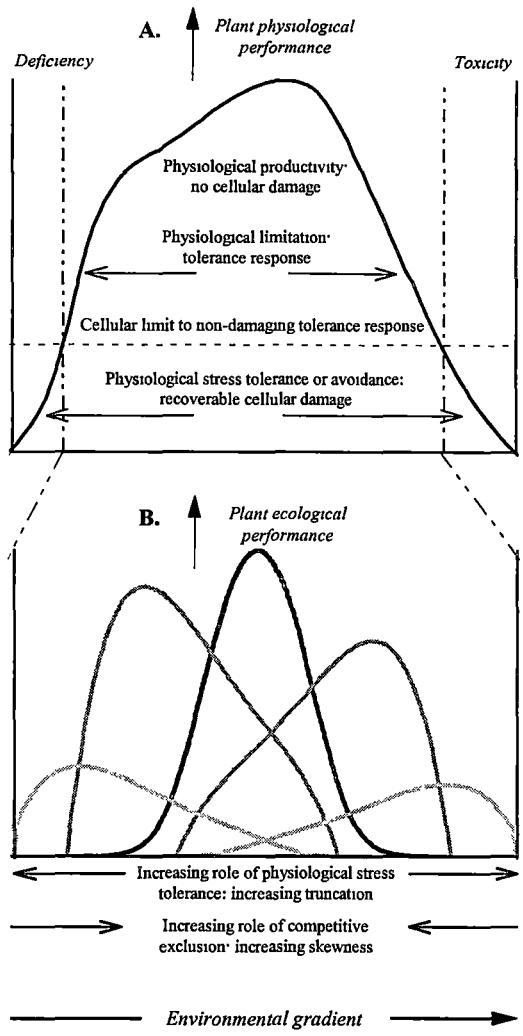


Figure 1.3 The nested arrangement between physiological and ecological responses (redrawn and reinterpreted from Austin 1990).

A. Schematic diagram of an asymmetric physiological response to variation in an environmental factor (e.g. temperature). The plant response is unlimited around the optimum, generally tolerant of limitations either side of the optimum, and under extreme conditions (deficiency or toxicity), cellular damage promotes mechanisms of stress tolerance or avoidance.

B. Schematic diagram of possible ecological responses to variation in an environmental factor (e.g. temperature). The ecological response approximates a family of possible curves nested within the physiological response. Apart from undefined historical factors, two mechanisms contribute to the shape of the ecological response — the outer envelope of the physiological response and negative biotic interactions, such as competition and predation.

Competitive ability, which enables a species to maintain an ecological optimum close to its physiological optimum, may be at the expense of its competitive ability toward the limits of its physiological tolerance (e.g. black response shape in Fig. 1.3B). Competitive exclusion that displaces the ecological response of a species either side of its physiological response, is limited by the physiological tolerance or avoidance traits of the species for environmental extremes of deficiency or toxicity. The displaced ecological response becomes increasingly tailed toward its physiological optimum and truncated toward its physiological limits (e.g. grey response shapes in Fig. 1.3B).

A displaced ecological response suggests the evolution of traits for physiological tolerance and avoidance that promote competitive ability in low resource environments (cf. Chapin *et al.* 1993a). Specialisation for traits of physiological tolerance and avoidance is usually at the cost of competitive ability in high resource environments. As a result, the shape of the ecological response may also have implications for the underlying shape of the physiological response, when this is not known.

1.6.1 The continuum concept and choice of environmental gradients

Interpretation of species' or community patterns in terms of the continuum concept requires that geographic distributions be considered as responses to environmental gradients that have a direct relationship to plant physiological processes (Austin & Smith 1989). To facilitate this approach, methods of 'direct gradient analysis' were developed (e.g. Whittaker 1967; Austin *et al.* 1984; Austin & Meyers 1996). However, the description of the ecological response of a species from a gradient analysis of distribution patterns may be confounded in several ways: (i) not all aspects of the environment may be taken into account; (ii) the necessary type and proximity of gradients to plant performance may not be fully considered; (iii) the scale between the response and explanatory variables may not be completely matched; and (iv) a systematic bias may exist in the data including errors in the recording of response or explanatory variables. As a result, considerable confusion over the role of direct gradient analysis and the appropriate definition of environmental gradients has arisen (e.g. Økland 1992). In addition, there appears to be some confusion between the levels of an environmental factor and its functional interpretation with respect to the productivity and stress effects on plant response (e.g. Grime 1977, 1988; Keddy 1990; Wisheu & Keddy 1992).

1.6.1.1 Indirect versus direct environmental gradients

In some extreme habitat types, indirect environmental conditions may appear to dominate the ecological response of species. This is probably due to their strong correlation with the actual gradients which control the physiological response. For example, water-table depth becomes a significant resource index for moisture and oxygen availability in Fennoscandian mires (e.g. Økland 1992), because of the high degree of correlation between the resource and its indirect measure. However, in forest habitats, other substrate factors (e.g. soil texture, structure and stone fraction) interact with water-table depth in their influence upon water supply rates. Consequently, there exists a lower correlation between the direct water resource supply gradient and the indirect depth-to-watertable gradient, consistent with the classification of water-table depth as an indirect environmental gradient (*sensu* Austin 1990). Therefore in some circumstances, indirect environmental gradients may provide adequate surrogates for the underlying direct effect when describing ecological processes, but the actual physiological basis of a plant response should not be overlooked if attempting an ecological interpretation of species' distribution patterns or extrapolating results to a new situation.

1.6.1.2 Disturbance gradients

Disturbance is a complex environmental factor including gradients in time and levels of one or more environmental factors. For example, in the context of the continuum concept, disturbance by fire may be viewed as comprising abiotic gradients in temperature, humidity and wind speed; a biotic gradient in accumulated fuel; and a temporal gradient in return-time of ignition (e.g. Clark *et al.* 1996; Keane *et al.* 1996a). Similarly, the creation and elimination of temporary pools of water (Southwood 1988) may be viewed as a temporal gradient in the return-frequency of a

gradient in free water and the shorter-term duration of its retention. Gap-dynamics in forest communities (e.g. Wayne & Bazzaz 1993a, b; Lertzman 1995; Sipe & Bazzaz 1994, 1995; Kuppers *et al.* 1996) may also be regarded as comprising gradients in the temporal probability of habitat-formation and the subsequent temporal-trace of resource availability. Other local disturbance gradients may involve the subtle effect of changing hydrological regimes on micro-habitat moisture availability (e.g. Wardell-Johnson & Horwitz 1996). Therefore, the consideration of disturbance factors in the gradient analysis of species' distributions requires their separation into component gradients related to the duration and intensity of physical and biological factors.

1.6.1.3 Stress and productivity gradients

The definition of environmental gradients as stress- or productivity-related confounds the two-sided characteristic of the plant response (e.g. Fig. 1.2). This interpretation clearly suggests that opposite ends of an environmental gradient, or indeed different environmental gradients, may actually be biological equivalents in terms of their effect upon plant response. This may be a reasonable reference for some ecological studies limited to a portion of an environmental gradient, such as the range from deficiency to optimum in water supply (rather than deficiency to water-logging). However, different plant physiological processes are likely to be involved in responses to deficient or toxic levels of the same environmental gradient (e.g. Grime 1979, 1988; Tilman 1982, 1988; Austin & Smith 1989).

The inherent continuity between stress and productivity functions of an environmental gradient has frequently been described by terms like 'stress/productivity' or by separate reference to gradients in stress or productivity (e.g. Sibly & Grime 1986). While such a classification of phenomena is at times convenient, it is also confusing (Taylor *et al.* 1990). Nevertheless, this has become a common way of classifying environment for the purpose of studying and interpreting vegetation patterns (e.g. Grime 1977, 1988; Keddy 1990; Wisheu & Keddy 1992). Analyses based on such *a priori* interpretation of environmental gradients may confound subsequent interpretation of plant responses.

I argue here that the classification of environmental gradients into their stress or productivity functions is to some extent arbitrary and removes the relationship between response and effect. It suggests that different responses may arise from the same underlying effect, which may not be the case. If we accept the general physiological characteristic of vegetation or plant response to variation in the environment (e.g. Fig. 1.3A), then it appears that some ecologists are describing environment in terms of the plant response, rather than in terms of the underlying levels of the environment effecting the response. This is not the same as when explicit physiological processes are used to simulate plant responses to environmental gradients, resulting in site index estimates related to productivity, or its inverse, stress (e.g. Battaglia & Sands 1997).

1.6.1.4 Collective vegetation properties as environmental gradients

Gradients in vegetation properties, including productivity (e.g. Keddy 1990) or compositional diversity (e.g. Økland 1992), have been used as a means of integrating environmental effects on plant response so that species can be ordered in one dimension. This represents a vegetational ordination (*sensu* Austin 1985) within which a particular plant characteristic (e.g. species, life-form, trait) is indirectly arrayed with respect to a gradient in vegetation performance (e.g. productivity, species richness, standing crop). This indirect gradient analysis (*sensu* Whittaker 1967) is a useful tool for emphasising competitive responses between species, and comparing hypotheses of niche arrangement. Species' responses are effectively standardised by the collective vegetation response as a surrogate for the productive capacity of a site. This may provide a useful summary of experimental and observed detail of plant performance in relation to environmental gradients. For example, the simulation of photosynthesis and respiration integrates our physiological understanding of plant responses to environmental gradients (e.g. Leemans and Solomon 1993; Wang & Polglase 1995). However, when productivity is considered as the organising principle in ecology, and this drives the collection of experimental information (e.g. Keddy 1990; Wisheu & Keddy 1992), then the subsequent data set is likely to be unsuitable for establishing vegetation responses to the component gradients of environment. This was pointed out by Austin and Gaywood (1994, p. 474):

> *Productivity is usually equated with standing biomass and is the result of an interaction between plants and their environment. There is therefore an element of circularity in using biomass as a surrogate for environment since changes in productivity may be due to a variety of environmental variables, such as temperature, light, water, nutrients, or carbon dioxide. There is no reason to believe that plants respond to gradients in the same way (Austin & Smith 1989). So, until it is shown to be otherwise, biomass or productivity is better considered as a collective property of the vegetation with potentially different responses to different gradients.*

Alternatively, collective vegetation properties can be considered as surrogate gradients for undefined environmental factors related to disturbance, competition and predation that complement information about the physical environment. For example, spatial variation in a species' response could be assumed to be adequately resolved by site-to-site variation in climate and substrate. Short-term temporal cycles of the environment at a site could also be partially resolved by including the magnitude of seasonal variations in climate within and between years. However, environmental variability which occurs over longer time periods, such as the return frequency of disturbance events which affect the successional development of the vegetation, may not be easily resolved.

The residual unexplained variation associated with a species' distribution, after considering the effects of spatial and temporal variability of climate and substrate, may be an indication of the influence of other factors (e.g. disturbance, competition, predation and biogeographic history) on the micro-habitat. Some of these factors may collectively influence the vegetation at a site, of which the species is one interacting component. Correlation between residual variation and

collective vegetation properties (e.g. the structure and composition of the vegetation) may therefore provide an indication of, or surrogate for, the relative importance of other physical and biological factors influencing the distribution of a species. Therefore, independent observations of vegetation height and cover as the biotic habitat context for a species could be used to distinguish differences between sites that would otherwise appear the same on the basis of their climate and substrate characteristics.

1.6.1.5 Which environmental gradients?

Climate and substrate interact with the genotype of a plant to determine the potential rate of resource supply and acquisition. Biotic interactions, through competition, facilitation and predation, influence the actual level of these resources available to a plant. An imbalance in internal resource levels thus leads to a trade-off in responses affecting the partitioning of biomass directed toward restoring the resource balance. The integrated economy of the plant response ensures a great deal of phenotypic flexibility in a fluctuating environment, and evolution encourages physiological specialisations in response to directional selection pressures over many regeneration cycles. Therefore, it can be expected that different sets of predictor variables will be applicable to different groups of species, and its is doubtful that the environmental hyper-space of a plant could be simplified for whole group of plants (Lawesson 1997).

To start with, environmental gradients appropriate to a theoretical study of vegetation patterns must be directly interpretable in terms of the physiological constraints on growth and survival of plants (Austin & Smith 1989). A continuous array of environmental gradients exists from distal to proximal in their influence upon plant physiological processes, and therefore, growth. Indicative variables from each of the primary factors — water, temperature, nutrients and light — that describe the spatial and temporal components of a plant's physical and biological environment would be needed. The closer the physiological relationship, the clearer the interpretation of physiological and ecological processes from an analysis of species' distributions. However, in ecological studies, the proximity of this micro-habitat environment can rarely be matched to the response. Nevertheless, surrogates which reasonably indicate the levels of a resource available to a plant could be used as gradients to be correlated with a response.

While physiological considerations guide the actual choice of factors to be correlated with a species' distribution, practical considerations determine the scale and availability of data on environmental gradients. For example, gradients in leaf and soil temperatures represent direct influences upon rates of plant responses for photosynthesis and respiration. But these cannot be easily measured or accurately modelled (e.g. Benecke & van der Ploeg 1981; Stathers *et al.* 1985; Waelbroeck 1993; Hanson *et al.* 1993; Callaway & King 1996; Diemer 1996). The less proximal estimates of atmospheric temperatures from climate models provide an approximation to the environmental gradients that have reasonably direct physiological consequences for plant responses (e.g. Toy *et al.* 1978; Agren *et al.* 1991; Bonan & van Cleve 1992; Franko *et al.* 1995).

Similarly, estimates of root-zone soil moisture availability and leaf or canopy boundary layer humidities are proximal estimates of micro-meteorological influences upon plant water balance that are rarely practical to measure for widespread studies (e.g. Rundel & Jarrell 1989; Shuttleworth 1989; Ragab 1995). However, process models of landscape water balance, linking climate models and soil properties (e.g. Hatton *et al.* 1993; Moore *et al.* 1993a; Pierce *et al.* 1993; Reiche 1994; Lathrop *et al.* 1995; Fitz *et al.* 1996) provide an indirect mechanism for approximating the water environment directly experienced by the plant. This applies also to remote sensing and terrain models (e.g. Running *et al.* 1989; Moore *et al.* 1993b; Zheng *et al.* 1996).

Light regimes are also readily estimated from physical process models linking the effects of location, topography, cloud cover and even vegetation structure (e.g. Iqbal 1983; Bonan 1989; Hook & McClendon 1992; Chen *et al.* 1993; Dubayah 1994; Forsythe *et al.* 1995; Clark *et al.* 1996; Yin 1996). Variables for solar radiation can be used as constraints on photosynthesis (e.g. Runyon *et al.* 1994; Leuning 1995; Dewar 1996b; Sands 1996). However, soil nutrient status is less readily quantified and modelled (e.g. Binkley and Vitousek 1989; Landsberg *et al.* 1991; Oja *et al.* 1995), and even if it could be quantified and modelled, there is no universal relationship between soil and plant nutrient status (Schulze *et al.* 1994).

In general, environmental gradients for the study of plant distributions are likely to be derived from the mean annual and seasonal extremes of climatic variables, as indices of their continuous variation (e.g. Lenihan 1993; Lawesson 1997). For example, meteorological observers of climate generally record two daily measures for temperature — minimum and maximum — which reflect overnight and daytime extremes and with some assumptions, enable diurnal variation to be interpolated. These day and night extremes of temperature are likely to have different effects on physiological processes. Therefore, maximum and minimum temperatures may provide an index of average plant response weighted by different day-time or night-time physiological processes. The relative significance of each to plant distribution patterns may indicate the nature of trade-offs in physiological responses to inter- and intra-annual environmental variability. Mean daily temperatures, however, assume that day- and night-time extremes have the same degrees of physiological influence, which may not be the case for all species or for all types of environment.

Similar interactions with physiological responses are likely to be associated with climatic indices for water availability. Meteorological observations for rainfall, evaporation and rain days represent components of water balance which, with substrate characteristics, could provide an indication of the potential seasonal variation in moisture regime for plant function. Scalars for nutrient status could also be derived from substrate characteristics or parent rock type (e.g. Ryan *et al.* 1995; Nix *et al.* 1992). However, in the absence of substrate descriptions (e.g. soil texture, soil depth, stone fraction), the nature of the interactions between the soil, the plant and the atmosphere must be assumed.

Additional surrogates for undefined environmental factors related to disturbance and other such landscape processes, or to competition and predation, could be defined from the habitat context of a species. Estimates of vegetation height and cover, or other collective community properties, are commonly recorded or mapped for inventory purposes (e.g. Hopkins & Graham 1981; Specht 1981; York *et al.* 1991; NSW National Parks and Wildlife Service 1995; Sulikowski 1995). Future advances in physical process models and calibration of remote sensing will improve the proximity and estimation of seasonal and annual levels of climatic gradients, levels of vegetation cover, terrain and soil characteristics or other physiographic features for ecological study (e.g. Ustin *et al.* 1991; Moore *et al.* 1993a, b; Running *et al.* 1994, 1996; Peterson & Waring 1994; Franklin 1995; Robitaille & Saucier 1996; Shao *et al.* 1996; Guenni 1997). Experimental calibration of remote-sensed data with improvements in climate modelling may also enable estimation of some more elusive environmental factors, including site nutrient status and water regime (e.g. Running *et al.* 1996; Nunez *et al.* 1996).

Climate indices therefore provide estimates for spatial variation of landscape processes at a given resolution (e.g. McMahon *et al.* 1996). The long-term norms of intra-annual temporal variation in the environment are also approximated. These are aspects of the environment experienced by a plant species. Given these constraints on meteorological information (local monitoring of climate rarely extends beyond a few decades), a minimum of three temporal indicators of annual and seasonal variability may be appropriate to consider for each climatic gradient as possible explanatory variables for species' distributions.

For the purpose of realised niche studies, Austin and co-workers (e.g. Austin 1980; Austin *et al.* 1984, 1990; Austin 1985, 1986, 1987, 1991a; Austin & Meyers 1996) proposed that more work was needed to define environmental factors which are as proximal as possible to the physiological response of plants. Apart from improvements in the resolution of process models and standardisation of remote-sensed data, the main factors which could be further refined for direct gradient analysis studies are water supply gradients, gradients in leaf and soil temperature, relative humidity, light quality and gradients in nutrient availability or rock chemistry. In addition, indices that represent interactions between these physical processes, and which have a direct effect on plant responses, need to be determined.

1.7 A framework for interpreting species' distribution patterns

The plant physiological response may have biochemical, anatomical and morphological implications for growth, tolerance, avoidance and defence, or developmental cues relevant to the life-cycle stage and the nature of the biotic or physical environmental gradient. The combinations of these response characteristics are finite, and recognisable patterns are evident as plant functional types. The ecological response of a species thus depends upon its functional type and the set of physiological capabilities reflected in the cumulative responses of its individuals to internal resource imbalances imposed by biotic and abiotic components of their environment.

Part of this genetic capability is the degree of phenotypic plasticity which enables a plant to optimise its physiological response toward directional variation in its environment, whether this is optimal or limiting. Observations of a plant's ecological response are therefore a direct reflection of the proximal effects of its environment that determine its physiological response. Although this proximal environment can rarely be measured in ecological studies, norms of reaction suggest that the plant response will be an approximate match to the stochastic variation of its evolutionary habitat.

Species' distributions can be predicted from the empirical correlation of relative performance with environmental gradients. Interpreting these predictions in terms of the continuum concept requires that responses be viewed relative to one or two environmental gradients, so that non-linear shapes and skewness become apparent. In the absence of specific population or genotype information, the characteristics of each response thus displayed could be related to known functional types and individual plant traits (e.g. Table 1.1). Where the nature of a species' physiological responses are known from experimental studies, these could be compared with the ecological response to assess the degree of 'competitive decrease' or shifts in relative position of optima along an environmental gradient (e.g. Fresco 1982; Malanson *et al.* 1992). The flexible definition of plant functional types from fundamental physiological and life-history processes also provides opportunities for interpreting a functional basis for species' ecological responses (e.g. Rutherford *et al.* 1995; Kelly 1996; Skarpe 1996).

A gradient model in plant traits is therefore developed from the concept of integrated plant physiological processes that lead to trade-off responses (Chapin 1991; Chapin *et al.* 1993a) and the continuum concept of species' distributions (Austin & Smith 1989). The objective of this model is to provide a process-orientated framework for interpreting the correlation between plant distributions and environment. For this purpose, species are viewed as super-genotypes: the patterns of response to environmental gradients as defined by the geographic distribution of populations are taken as indicative of the phenotypic response of a plant genotype. All plant species are viewed as genotypes of a super-species (the plant kingdom), and biotic or abiotic environmental gradients are referred to as varying from deficient to toxic, relative to all possible levels of performance.

Early classifications of species' functional types were based upon only a few environmental conditions that relate to levels and types of competition, stress, disturbance or productivity (e.g. competitor, stress tolerator and ruderal plant strategies of Grime 1977). These classifications could not completely account for the potential combination of physiological traits leading to different plant strategies (e.g. Southwood 1988; Taylor *et al.* 1990). An alternative approach to a classification of functional types was utilised by Smith and Huston (1989) based on fundamental niche constraints affecting the ability of a plant to grow under simultaneous limitation in two or more resources, demonstrated for gradients in light and water availability. Their model is defined from trade-offs in the ability of plants to adjust physiological and life-history traits to tolerate

limited resources, and adequately reproduces physiological as well as ecological responses. A similar trade-off model was developed by Oksanen and Ranta (1992) using grazing intensity (biotic gradient) and grazing frequency (temporal gradient) which interact with overall environmental fertility (abiotic gradients).

Similar premises about the consequences for growth and fecundity of the simultaneous use of other plant resources, (e.g. light and nutrients, nutrients and water, and the effects of temperature) have also been developed (e.g. Chapin *et al.* 1987, 1993a; Ellner 1987; Chapin 1991; Barton 1993; Sultan & Bazzaz 1993a-c; Huston & DeAngelis 1994; Lavorel & Chesson 1995; Sekimura 1995; Harrington *et al.* 1996; Kuppers *et al.* 1996). These concepts provide the flexibility necessary to simulate the range of putative functional types which may arise from different combinations of individual elements of physiological, morphological and demographic response to the driving forces of environmental change (e.g. Box 1996; Bugmann 1996a; Steffen 1996). This trade-off approach provides a framework and a set of rules that define the mix of functional types for predicting the transient response of vegetation within a particular ecosystem (e.g. Bugmann 1996a; Chapin *et al.* 1996a; Condit *et al.* 1996b; Steffen *et al.* 1996a). It also provides a mechanism for interpreting the combination of plant traits that may be selected by particular types of environment, and the types of competitive interactions which may prevail (e.g. Huston & DeAngelis 1994, Grace 1995).

The ecological response of a super-genotype to an environmental gradient may be viewed as an indication of a phenotypic trait (physiological, morphological, or life-history) which reflects a strategy for resource exploitation, stress tolerance or stress avoidance (Fig. 1.4). This is comparable to traits defined as competitor, stress tolerator and ruderal by Grime (1977), but allows for the possibility that each species may express a different trait in relation to a different environmental gradient, and these may vary with life-form stage. Dormant life-history stages (e.g. pollen, seed, deciduousness) may be viewed as ruderal traits of stress avoidance. For example, where drought is characteristic of a particular environment, plants may adopt avoidance strategies of deciduousness, balanced by mechanisms to maximise productivity when water is seasonally available (e.g. Rundel *et al.* 1995). Similarly, different plant organs, such as epicormic buds and lignotubers in eucalypts, may be viewed as strategies to tolerate or avoid occasional environmental extremes of fire and drought (e.g. Canadell & Zedler 1994; Florence 1996). The ecological response of a species is thus considered to be the cumulative outcome of these underlying traits that promote the survival of individuals and populations throughout its range. For example, the effects of a fire event on a forested landscape will depend on the intensity of fire, the sensitivity of species, their recovery strategies by vegetative or seed responses, and the interaction between component species during the regeneration phase (Florence 1996).

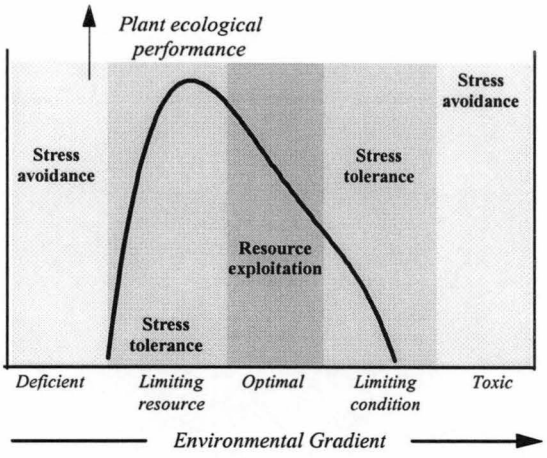


Figure 1.4 Arbitrary classification of plant strategies representing fundamental niche responses of stress avoidance, stress tolerance or resource exploitation according to levels of an environmental gradient. The form and position of an ecological response along this gradient shows greatest performance associated with limiting levels of the resource. Plant traits dominated by stress tolerance are expected to be associated with optimisation for a low resource environment.

Viewed with respect to environmental gradients of water availability and light intensity, varying between extremes of deficiency and toxicity, the classification of functional types defined by Smith and Huston (1989) explains only one quarter of the possible types of plant ecological response (Fig. 1.5). The intersection between the two methods of display suggests that a curvilinear delineation between existent and non-existent functional types may be appropriate. In addition, combinations of environmental conditions which are indicated by Smith and Huston (1989) as being the limit to plant survival may support plant adaptive traits of avoidance.

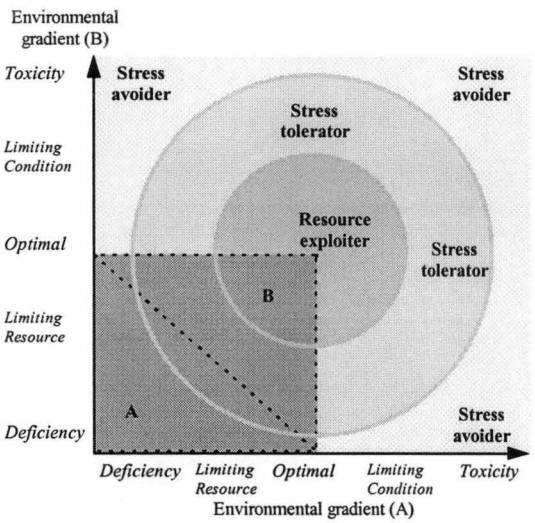


Figure 1.5 Gradient interpretation of adaptive plant traits from ecological responses. Environmental gradients vary from deficiency to toxicity in their proximal effect upon plant physiological responses. The region of plant response considered by Smith and Huston (1989) is defined by the *square of dotted lines* in which the diagonal defines their limit to the combined tolerance of shade and drought. Plant responses in region **A** represent the survival limit in the presence of low levels of water and light, and in region **B** a continuum of adaptive plant responses may be arbitrarily characterised into discrete functional types (see Smith & Huston 1989).

Although whole-plant functional types of avoidance are impossible, some plants have specialised mechanisms of restoration physiology to recover from otherwise lethal conditions (e.g. Pugnaire *et al.* 1996). Dormancy may also ensure propagule survival during dispersal or periods of environmental extremes (e.g. McDonald *et al.* 1996; Pake & Venable 1996). Other forms of opportunistic life-history strategies enable plants to move onto sites subject to recurrent disturbance, soon after the disturbance (e.g. Esler & Cowling 1995). Therefore, different sets of environmental conditions may be limiting or lethal at different stages of a species' life cycle. Different adaptive traits for the seed, seedling and adult phase will subsequently have developed. Species' responses to environmental gradients will then depend upon the interaction between the

physiological constraints set by their genotypes, their life cycle stages, the type of gradient involved, its variation over time and its interaction with other factors. Numerous dimensions of environment will thus need to be considered to account for the array of plant responses, when describing the realised niche.

The realised niche of a species may be defined by the intersection of the position of the optimum response to each component gradient for constituent genotypes (Fig. 1.6). Individual life-history traits influence the form of the whole-plant ecological response with respect to particular environmental gradients. The overlapping optimum response region defines where successful reproduction ensures the self-perpetuation of populations of the species. The species does not occur in environments that correspond to tails of the respective responses. Growth is insufficient to support reproduction under these conditions (cf. Fig. 1.1). These hypothetical tails of response represent the regions of fundamental constraints on whole plant survival due to trade-offs between tolerance and productivity (e.g. Chapin *et al.* 1993a), and the consequent limits to tolerance of two or more stress-related conditions (e.g. Smith & Huston 1989).

An example of interpreting the hypothetical ecological responses of species illustrated in Fig. 1.6 can be taken by considering the two environmental gradients as a gradient in the magnitude of temperature from high to low, and a gradient in temperature variability from equable to variable. Experimental studies of responses to temperature suggest that respiratory properties are the characteristics of plants most closely adapted to the environment so as to optimise survival and reproduction (Hansen *et al.* 1996). They conclude that genotypes can be placed into one of two major categories: plants that grow best in cool, variable temperature climates and plants that grow best in warm, constant temperature climates. Relative values of the temperature coefficients for respiration can be correlated with each of these groups. Therefore, the patterns of optimum and range in ecological response to temperature, demonstrated from correlative models of a species' distribution, can be related to one of these habitat categories and associated with specific adaptive traits.

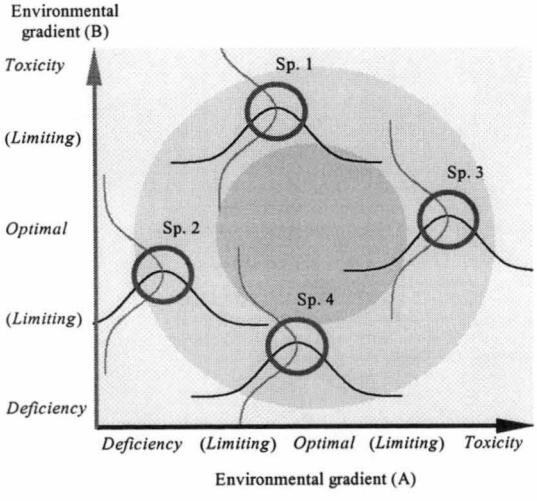


Figure 1.6 Demonstrating a continuum in adaptive plant traits relative to two environmental gradients. Two dimensions of a realised niche are shown for four hypothetical species by their overlain, univariate responses to component environmental gradients. The intersection of the position of the regions of greatest performance represent conditions suitable for reproduction. The tailed regions of each univariate response represent the positions of multiple stress and survival, but poor reproductive success. Each species occupies different core positions for given combinations of environmental gradients, representing different specialisations or adaptive traits. Symmetric responses are shown, but asymmetric responses are just as likely.

Each gradient will impose a selective pressure, and the position of an optimum response by a species or genotype indicates that an adaptive solution for survival has been achieved. In addition, species may exist in the same set of environments for different reasons. There may be more than one combination of stable adaptive traits which supports a similar optimum response, despite physiological constraints on their design (Southwood 1988), or species may have common evolutionary histories (Chapin *et al.* 1993a).

The flexibility of this gradient model of plant traits is due to few assumptions and its consistency with a 'bottom-up' approach based on individual plant responses, combining physiological and life-history attributes associated with proximal environmental gradients. Smith & Huston (1989) demonstrated how different functional strategies of resource use allow plants to fill different ecological roles under different resource conditions, and concluded that succession and zonation result from the same basic processes. The same basic theoretical concepts that underlie their trade-off model of vegetation dynamics also apply to this gradient interpretation of adaptive plant traits (Figs. 1.4-1.6).

In many ecological studies, the shape of the underlying physiological response for coexisting species is not known, and each environmental gradient is rarely completely proximal to physiological processes. Observational studies are also confounded by inconsistencies in scale, the differing chronosequences of succession at every point along a spatial gradient, and inadequate empirical or process information about the underlying shape of the factorial responses. However, theories about the fundamental nature of the individual-based response, and simulation models of these responses provide an indication as to the type of ecological traits which may be expected. An understanding of these component processes and their ecological roles may therefore facilitate physiological and life history trait explanations for response patterns obtained from empirical correlations between plant distributions and their environments.

1.8 Empirical methods for predicting species' distributions

Systematic observations of species' distributions have been accumulating over the past 50 years or more, culminating in publication of floristic classifications (e.g. Duncan & Brown 1985; Kirkpatrick *et al.* 1988), regional vegetation distribution maps (e.g. Kirkpatrick & Dickinson 1984) and flora atlases (e.g. Brown *et al.* 1983; Garrett 1996). The central compilation of these vegetation-environment inventories provides a convenient basis for reviewing the state of our knowledge by developing predictive models of species' performance (e.g. Kittel *et al.* 1995; Austin *et al.* 1996).

There have been two main approaches to prediction of plant ecological performance.

The first is based around the simulation of performance following the experimental description of physiological and ecological responses, including demographic behaviour (e.g. Acevedo *et al.* 1996; Shugart & Smith 1996). This has been discussed in previous sections in the context of

experimental and theoretical concepts of plant physiological and ecological responses. The extrapolation of these individual-based models to different sites, or scenarios of environmental change, can be performed using mapped inventory data to test and evaluate model results (e.g. Foley *et al.* 1996; Haxeltine & Prentice 1996a). Alternatively, inventory information can be used to iteratively develop bioclimatic indices for predicting plant distributions (e.g. Shao & Halpin 1995; Sykes *et al.* 1996).

The second approach is parametric correlation, using statistical methods of regression to relate plant response to environment (e.g. Austin *et al.* 1984, 1990). This method requires little prior knowledge about the underlying ecological traits of species, or the exact nature of their physiological performance. Therefore, models can be developed before detailed experimental information is available about the nature of species' responses or their types of demographic behaviour. However, interpretation of the results of such models requires that putative plant traits and functional types be inferred from the form of species' responses to individual environmental gradients (e.g. Fig. 1.6). In some cases, direct reference to the shape of responses derived from experimental studies will facilitate interpretation (e.g. Prentice *et al.* 1991; Chapin *et al.* 1993b). Alternatively, the results of statistical correlations could be used to guide and focus experimental designs.

Whichever the method applied, the purpose of analysis needs to be clearly defined. The purpose will influence the methods used to acquire the data, and the composition of the sample. It determines the observation scale and the degree of replication needed to achieve the necessary level of precision in prediction. A statement of the consistency of sampling with underlying ecological and statistical assumptions is also needed before analysis. Then, subsequent ecological interpretation in terms of the physiological function of species may be possible. Statistical methods of correlation are therefore reviewed here, given the purpose to predict plant distribution patterns from inventory information of their occurrences and environment, with little experimental knowledge of their performance.

1.8.1 Statistical correlation of species' distributions and environment

The objective of statistical correlation analyses by regression, is to determine which environmental factors are significantly related to the occurrence of a species. The form of the response variable in plant inventory data determines the statistical approach to prediction. Where reliable information is available for species' absences, in addition to presence observations, then the response is binary and the underlying statistical distribution is binomial. Such non-normal statistical distributions can be analysed through generalised regression procedures, known as generalised linear modelling (e.g. McCullagh & Nelder 1989; Hosmer & Lemeshow 1989; Buja *et al.* 1989). Where inventory data provides reliable information for the presence of a species only, and without information on relative performance, then different statistical methods of analysis are needed.

1.8.1.1 Analyses based on presence-only data

Presence-only data for the occurrence of a species has been analysed using a bioclimatic domain approach (e.g. Backhouse & Burgess 1995; Hughes *et al.* 1996). For example, the BIOCLIM method (Nix 1986; Busby 1986, 1991; McMahon *et al.* 1996) hypothesises that the potential range of a species is the totality of sites with a climate identical or similar to the climate at one or more sites at which the presence of the species has been recorded (e.g. Booth & Jovanovic 1988; Neave & Norton 1991; Lindenmayer *et al.* 1996). Climate surfaces, defined from the interpolation of meteorology data on a geographic grid (Hutchinson 1989), provide the mechanism for defining the climate profile for sites that record the presence of a species.

These climate profiles comprise specified indicator variables for mean, minimum, maximum, and seasonality of, temperature, rainfall, evaporation and other climate factors that are considered important correlates of species' distribution patterns. The potential distribution of a species is predicted by matching its climate profile with similar sites on a geographic grid. Prediction error is usually reduced by arbitrarily constraining prediction regions to percentile limits of the climate profile. For example, the core distribution of a species may be explicitly defined as the 90th percentile range (i.e. 5th to 95th percentile values) of the climate profile, and marginal distributions are those which fall outside this domain (Busby 1991). The resolution of the geographic grid, the choice of climatic factors, and the quality of the sample data influence the accuracy of prediction.

Elaborations of the bioclimatic domain approach for estimating species' potential ranges have since developed. For example, Walker & Cocks (1991) overcame some of the problems of misclassification in BIOCLIM by using a convex hull model. Carpenter *et al.* (1993) further extended these methods by estimating potential distributions using a range-standardised point-to-point similarity metric, with no discrete boundary for the environmental envelope. Shao and Halpin (1995) found that statistical censoring of the climatic extremes of data removed potential exaggeration of the core range of a species derived from mapped occurrences and significantly increased the predictive ability of their elliptical distribution model.

Alternatively, the results of mechanistic modelling that define plant performance parameters, such as persistence or productivity, could be coupled with climatic surfaces to define threshold values of a species' bioclimatic domain (e.g. Hill 1996).

In lieu of comprehensive information, these environmental domain methods provide a mechanism for defining the potential distribution of a species (Boston & Stockwell 1995). However the resulting predictions represent conditional probabilities only: conditional on the presence of a species. In terms of the ecological response of a species, such analyses exclude the possibility that an absence is a result of lack of sampling rather than a real absence.

For example, take the hypothetical case of a widespread species with occurrences ranging from lowland to highland habitats (Fig. 1.7). The landscape categories for the presences suggest a

strong tendency to occupy lowland habitats. However, if the declining availability of highland habitats is taken into account, then the probability distribution suggests no particular preference. The ecological interpretation depends upon the type of analysis, i.e. presence-only data (conditional frequencies) versus presence/absence data (probability of occurrence).

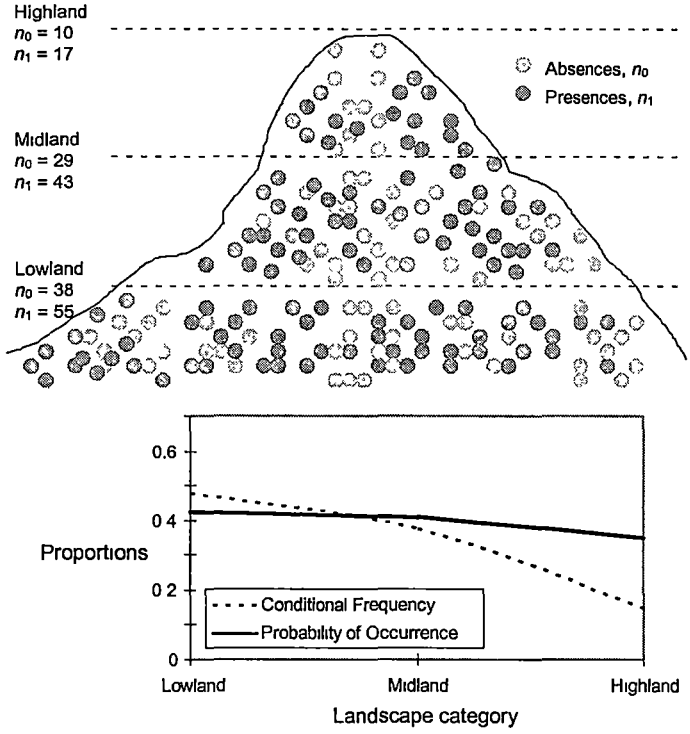


Figure 1.7 An imagined distribution of a species in three landscape categories for an imagined mountain profile. Locations for random samples of presence and absence observations for the species are shown.

Conditional frequencies are calculated as the proportion of presences in each category relative to the total number of presences across all categories.

Probability of occurrence is calculated as the proportion of presences relative to the total number of presences and absences in that category.

1.8.1.2 Analyses based on presence and absence data

Floristic inventories are ideal sources of absence information, especially for prominent species, such as mature forest trees. Statistical methods for the analysis of binary responses enhanced the regression analysis of these plant distribution data (e.g. Cunningham *et al.* 1981; Austin *et al.* 1984, 1990, 1994; Bartlein *et al.* 1986; Westman 1991; Yee & Mitchell 1991; Leathwick & Mitchell 1992; Lenihan 1993; Brown 1994; de Swart *et al.* 1994; Stahle & Chaney 1994; Franklin 1995; Leathwick 1995; Austin & Meyers 1996; Wu & Huffer 1997). Previous analyses were restricted by the inability to include reliable information about unoccupied types of habitat (e.g. Swan 1970; Austin 1971). Information about the inherent presence/absence response associated with species' distribution patterns enables the conditions associated with unsuitable habitat to be considered along with suitable habitat in a statistical analysis.

Logistic regression is the most common application of generalised linear modelling when the response is binary or dichotomous (Hosmer & Lemeshow 1989). Other possible link functions for the transformation of the response variable are the probit (inverse Normal function), the complementary log-log and the log-log function (McCullagh & Nelder 1989). The logistic (also known as logit) and the complementary log-log function have similar properties over the probability interval 0.1 to 0.9, and are difficult to discriminate on the grounds of goodness-of-fit

(McCullagh & Nelder 1989). This is the usual range in predicted probabilities of occurrence from logistic regression analysis of plant distribution models (e.g. Austin *et al.* 1990). Since much of the current plant ecological literature has been developed on the logistic response, and there is no discernible difference between the performance of alternative error functions, there seems little point in digression from logistic regression.

One important property of the logistic function which makes it appropriate to the analysis of plant distribution data is that differences on the logistic scale can be estimated regardless of whether the data are sampled *prospectively* or *retrospectively* (McCullagh & Nelder 1989). Plant distribution data is retrospective because the influence of the environment upon the performance of a species at any one site is observed subsequent to the selection pressures and adaptive responses that led to its population occurrence. In retrospective data, the intercept is an 'incidental' parameter: necessary for the logistic regression model to make sense physically, but largely irrelevant to the conclusions which result from the analysis (McCullagh & Nelder 1989).

1.8.1.3 Minimum common information from plant inventory data

Inventory data of vegetation composition from regional and local land surveys is a cost-effective source of information that can be readily compiled to form large sets of ecological data (Austin 1991a). These ecological data generally comprise the minimum information necessary to broadly map distribution patterns of plant species (e.g. Brown *et al.* 1983; Maslin & Pedley 1988; Saetersdal 1994; Garrett 1996). A subset of these data contain additional habitat information for a statistical analysis of vegetation and environment relationships.

In addition to records of presence and absence, associated climate and substrate information are essential to an analysis of species' distribution patterns. Further site information related to soil characteristics, terrain descriptions, disturbance regime and attributes for vegetation structure, would be useful for predictive modelling, but are rarely consistently recorded between sources. Other response variables for species' relative performance, would enable a wider range of questions related to habitat selection and competition (e.g. Shugart & Urban 1989; Bazzaz 1991) to be addressed from plant distribution patterns (e.g. Lenihan 1993).

1.8.2 Statistical assumptions of regression methods

The application of regression analysis to the study of plant distribution data requires that the approach be consistent with the statistical assumptions of the method. The use of statistics assumes that the data were randomly sampled such that they represent the characteristics of the population being studied (e.g. species' geographic distribution patterns). The method also assumes that each of the samples are independent and not autocorrelated, that is, that the values for the response of one sample do not influence the response of an adjacent sample. Plant distribution data can be spatially autocorrelated, and experimentally monitored data can be temporally autocorrelated. Similarly, the use of logistic regression methods assumes that the data are binomially distributed and that the logit is the appropriate transformation. Checks on the

statistical assumptions of the sample data and the model are a necessary feature of exploratory analyses and regression modelling.

1.8.2.1 Sampling considerations

Predictions based on presence only, or presence and absence' responses will be influenced by any sampling bias that exists in the data. This is because statistical methods inherently try to find the best fit between the response and explanatory variables. Therefore, the sampling adequacy of either type of data with respect to the purpose of prediction needs to be assessed. In addition, because the pattern of presence and absence information determines the response in logistic regression, it is important that these be representative of their configuration in the landscape. Bias in the balance of presence and absence records may lead to spurious ecological correlations in subsequent analyses, and confound interpretation. A representative sample which is based on the theoretical relationship between species' distributions and environment is therefore an essential pre-requisite for predictive modelling.

Random sampling and representativeness

Statistical analysis techniques generally assume that a random method of sampling was used to acquire the data, and that this inherently leads to a representative sample. In observational studies, random sampling techniques are designed to minimise the effects of personal judgment in the choice of sample location and to maximise the statistical validity of the recorded information (Pardo 1994). However for regional ecological survey where access is limited, stratified random sampling techniques are preferred as a cost-effective means of obtaining a representative sample of the vegetation (e.g. Noy-Meir 1971; Bunce & Shaw 1973; Austin & Basinski 1978; Orloci & Stanek 1979). A carefully designed method of stratified random sampling returns a representative sample with equivalent statistical properties to simple random sampling (Pardo 1994).

For the purpose of predicting vegetation-environment relationships, methods of stratified random sampling for systematic ecological survey aim to classify the landscape according to features of the environment that are believed to influence vegetation patterns. Replicate samples of the vegetation are obtained according to their random placement within environmentally defined units of the landscape. In remote situations, an explicit bias for accessible locations may also be included whilst maintaining the emphasis of geographic and environmental representativeness (the 'gradsect' approach - Gillison 1984; Gillison & Brewer 1985). These methods were formally applied and tested by Austin & Heyligers (1989, 1991) and further discussed by Margules & Austin (1994). The concept of geographically and environmentally representative stratification is now widely used in the design of regional biophysical surveys to obtain an indicative sample of habitat for the predictive study of vegetation (e.g. McKenzie *et al.* 1991; Ludwig & Tongway 1995; Fahey & Crow 1995) and fauna (e.g. Braithwaite *et al.* 1989; Ferrier & Smith 1990; Mills *et al.* 1996; Neave *et al.* 1996b; Jones & Rose 1996).

Systematic sampling is subsequently assumed to have recovered a representative sample for the purpose of predictive analysis (e.g. Ferrier & Watson 1994, 1996). However, centrally compiled ecological surveys represent an *ad hoc* sample and, because of differences in sampling intensity or scale by different surveyors, these data are unlikely to be representative (Austin 1991a). A retrospective analysis of sampling adequacy is therefore needed before compiled ecological data are used to predict species' distribution patterns.

'Naughty-noughts' and sampling domains

Predictive analyses based on the correlation patterns between species' occurrence and factors of their habitat can be biased by sample data which include a large number of absence records beyond the species' environmental range (Austin & Meyers 1996). This phenomenon of analysis has been termed 'naughty-noughts' (Austin 1979; see also discussion in Mead 1971; Bartlein *et al.* 1986). It is particularly a problem with rare or uncommon species. A large number of trailing absence records relative to presence records along an environmental gradient reduces the precision of analysis and may result in species being predicted where they do not occur (Austin *et al.* 1990). Limits to prediction can be subjectively defined by setting upper and lower bounds for key environmental factors that include sufficient absence records to adequately define the form of response (e.g. Austin *et al.* 1990, 1994; Austin & Meyers 1996).

Appreciation of these environmental constraints can be further enhanced by setting the limits in a beta-function curve which analytically approximates a skewed or truncated response to temperature (e.g. Austin 1992; Austin *et al.* 1994). However, Oksanen (1997) shows that the beta-function is not suitable for testing the shape of species' response curves because the location of the optimum and skewness are correlated so that these parameters cannot be estimated independently. An alternative way of defining the sampling domain for presence and absence information may be to restrict the data to those from the potential geographic and environmental range of the species, with flexible approaches to modelling species' response shapes. An estimate of species' potential geographic ranges also provides a context for assessing representativeness of presence and absence responses.

1.8.2.2 Spatial autocorrelation in plant distribution data

Regression assumes that samples are not autocorrelated (i.e. that the error terms in a model are not correlated). Autocorrelation usually arises in data that are serially collected, either in space or time, resulting in values of a response for one sample influencing the response of an adjacent sample (McPherson 1990). Spatial autocorrelation (positively or negatively) is a common problem of plant distribution data because sites are spatially connected by environmental factors, such as climate, and therefore by plant response to climate (e.g. Wildi 1991; Legendre 1993; Smith 1994; Henebry 1995). Alternatively, the omission of one or several key variables can be a cause of the autocorrelation effect (Wetherill *et al.* 1986; Neter *et al.* 1989). Omitted variables are a likely cause of some instances of autocorrelation in plant distribution data, either because of measurement difficulties or unknown factors. For example, the geographic distribution of plant

species may be influenced by undefined spatial factors related to dispersal probabilities, but these can be rarely quantified for explicitly including as explanatory variables in regression relationships.

Autocorrelation is also exacerbated when samples for analysis are derived from contiguous or near-contiguous vegetation occurrences, as in the case of transect sampling or the acquisition of information from landscape maps (e.g. Smith 1994; Clark *et al.* 1996; Thomson *et al.* 1996; Wu & Huffer 1997). However, for forest vegetation samples of single-tree canopy characteristics, Clark *et al.* (1996) found that separations of more than 20 m were adequate to ensure sample independence. Forest inventory samples record vegetation response and population occurrence across larger areas (i.e. 0.1 to 0.3 ha plots). Scaling up then, sampling independence is likely to be achieved by separation distances greater than about 100 m. While some inventory surveys are based on wandering transects (e.g. Kirkpatrick 1981), central compilation of different surveys is likely to reduce the effect of autocorrelation between samples in the large data sets.

Autocorrelation has several consequences for regression analysis (Neter *et al.* 1989). For example, while the estimated regression coefficients are still unbiased, they no longer have the minimum variance properties and may be inefficient. The mean square error may seriously underestimate the variance of the error terms. The true standard deviation of the estimated regression coefficient maybe seriously under-estimated. Thus spatial autocorrelation in plant distribution data impairs our ability to perform standard statistical tests (Legendre 1993). Spatial autocorrelation in a regression model could be subjectively detected by plotting residuals by the dimensions for geographic location (i.e. latitude, longitude, altitude) (see Neter *et al.* 1989). Several formal tests also exist (e.g. see Neter *et al.* 1989; McPherson 1990). Legendre (1993), in particular, develops a specific strategy for dealing with spatial autocorrelation in ecological data.

Ferrier & Watson (1994) dealt with the potential problem of spatial autocorrelation in distribution data by including the variables for planar geographic space (latitude and longitude) as covariates in their predictive models. An alternative approach was devised by Smith (1994), who constructed neighbour variables as spatial autocorrelation functions when modelling data derived from mapped sources. Wu & Huffer (1997) dealt with a similar problem of spatial autocorrelation in mapped data sources. They developed an autologistic regression model which gave a more parsimonious solution and better overall fit in predictive models with climate factors, than did ordinary logistic regression which 'ignores the spatial correlation in the responses and attempts to explain all the variation using the climate covariates'. This method is suited to contiguous responses sourced from mapped information, but may not necessarily be applicable to plot samples that are the basis of plant inventory data.

Since autocorrelation may also arise due to omitted variables, and because plant inventory data will have minor instances of near-contiguous sampling, measures for explicitly detecting and dealing with the problem (e.g. Legendre 1993) may be an unnecessary distraction. The general

adherence to statistical sampling requirements of representation and adequate replication in plant inventory data may be a simpler way of addressing confounding effects due to spatial auto-correlation. Simple retrospective detection of spatial autocorrelation in plots of residuals by spatial variables may provide an indication of the magnitude of the effect, and thus suggest whether specific analyses may be needed.

1.8.2.3 Testing the choice of link function and regression diagnostics

The predictive application of the results of a statistical model depend upon the appropriateness of the assumptions which underlie the model. For example, extrapolation beyond the range of the observed values of the explanatory variables, in order to predict the probability of occurrence of a species in extreme environments, depends upon the robustness of the assumed model and particularly on the choice of the link function (McCullagh & Nelder 1989). For a correctly specified model, the goodness-of-fit statistic (e.g. Pearson, chi-squared or deviance) should be approximately equal to the degrees of freedom. In practice the ratio of these two values often exceeds unity, indicating over-dispersion of the data and problems in the specification of the model (McCullagh & Nelder 1989; Hosmer & Lemeshow 1989). This may be due to outliers in the data, the use of the wrong link function, omission of important terms, or the need to transform some of the predictors. Regression diagnostics provides a means of assessing these problems before applying methods to rescale the covariance matrix to correct over-dispersion.

Diagnostic statistics to identify influential observations and to quantify the effects on various aspects of the maximum likelihood fit were developed by Pregibon (1981) for binary response data (see also Hosmer & Lemeshow 1989). A systematic approach to the application of these methods is included in current statistical software packages (e.g. SAS Institute Inc. 1997).

1.8.2.4 Prediction error evaluation

The limitations and uncertainties associated with predictive models need to be dealt with explicitly, systematically and efficiently (Norton & Williams 1992). Methods for evaluating the predictive performance of models developed using logistic regression were reviewed by Ferrier and Pearce (1996) and Ferrier and Watson (1996). Their reviews largely derive from the development of validation techniques for logistic regression in other disciplines, such as weather forecasting, psychology and medicine (e.g. Hanley & McNeil 1982, 1983; Swets 1988; Hosmer & Lemeshow 1989; Miller *et al.* 1991; Metz *et al.* 1993; Van Houwelingen & Le Cessie 1990). Unbiased estimates of prediction error are needed to determine how the model might be applied to questions of land use or planning, and whether constraints on the use or interpretation of the model might be needed (Ferrier & Watson 1994).

Resubstitution is the simplest and most frequently used method of evaluating the fit between the model and the data. Resubstitution analyses involve the use of the data from which the model was derived as the basis for assessing the fit of the model. These methods generally form part of modern statistical software packages (e.g. SAS Institute Inc. 1997). However, Ferrier and

Watson (1996) advocate alternative techniques because resubstitution yields an inflated estimate of model performance (or deflated estimate of error rate). They concluded that a more rigorous approach to evaluating model performance is to validate predictions using an independent sample. This should be geographically and environmentally representative of the modelled response, or of similar quality to the sample from which the model was derived.

The evaluation techniques outlined by Ferrier and co-workers aim to determine the strengths and limitations of a particular model by assessing its calibration, discrimination and refinement. Model calibration describes the tendency to over- or under-predict the occurrence of a species at a site. Model refinement is related to discrimination and describes the spread of predicted probabilities. A well-refined model will produce predicted probabilities close to zero and one (representing either a presence or an absence, rather than intermediate levels of response). The discrimination capacity of a model describes the ability to distinguish correctly between occupied and unoccupied sites, and the extent to which actual presences have higher predicted probabilities than actual absences. Knowledge of the relative calibration, refinement and discrimination ability of a model enables the precise use for which the model is suited to be determined (Ferrier & Pearce 1996; Ferrier & Watson 1996).

For example, errors of prediction can be assessed by considering the cross-classification of the observed and predicted responses as binary responses defined for specified threshold values (e.g. see classification table in PROC LOGISTIC, SAS Institute Inc. 1997). For a selected threshold value (e.g. 0.5), misclassification of an occurrence as an absence when it is in fact a presence may be a more important consideration in terms of assessing the risks of extinction to a species, than the reverse (e.g. McCarthy *et al.* 1996). Over-prediction of a presence may also be unhelpful in land use planning, leading to increased costs associated with field checking if decisions are based on a precautionary principle (e.g. Moir & Mowrer 1995; Dovers *et al.* 1996). The selection of the threshold probability and the utility of a model may therefore depend upon its overall ability to correctly predict either a presence or an absence, and the relative importance of either to the purpose in mind.

1.8.2.5 Transformation of explanatory variables and polynomial terms

Statistical model building requires that potential non-linear relationships between response and explanatory variables, that might be expected from ecological theory, be taken into account (Austin 1987). The functional shape of the relationships between the species' response and each explanatory variable is also of fundamental interest to plant ecology (e.g. Huisman *et al.* 1994; Austin *et al.* 1984, 1994; Austin & Gaywood 1994; Austin & Meyers 1996; Oksanen 1997). Whether species' ecological responses are Gaussian in form, or skewed in one or other direction, or complex in shape with more than one optimum position, has been widely discussed (e.g. Ellenberg in Mueller-Dombois & Ellenberg 1974; Gauch & Whittaker 1972; Gauch & Chase 1974; Austin & Gaywood 1994). Particular response shapes could also be a characteristic response of functional types to a particular environmental gradient, such as mean annual

temperature (e.g. Austin 1992). The form of a species' ecological response provides a basis for inferring optimum habitats, specialist or generalist phenotypic tolerances and the symmetry of putative competitive interactions (e.g. Smith & Huston 1989; Malanson 1997). However, prior to analysis, species' response shapes are generally not known because of the multiplicity of factors involved.

Since the actual form of a species' distribution response is not known, options for the nature of this relationship need to be considered as candidates in ecological models. Typically, the curvature of the response has been modelled through polynomials, and more recently a beta function has recently been introduced (e.g. Austin 1992; Austin *et al.* 1994; although see Oksanen 1997). Other forms could involve exponentiation or some other transformation (e.g. square-root, cube-root, natural log) of the explanatory variables; or an experimentally-determined non-linear function (e.g. shapes with different properties and parameters, see Ratkowsky 1990). More explicitly, Huisman *et al.* (1993) describe a set of non-linear models for hierarchical testing in species' response analysis. Previously it was believed that non-linear specification of responses could not be applied with generalised linear modelling (e.g. Huisman et al. 1993). However, this is not the case. Such non-linear functions for explanatory variables can be specified in generalised linear models (e.g. Oksanen *et al.* 1990; Oksanen 1997; D. Ratkowsky personal communication, January 1997), so there is no statistical reason for avoiding their use (e.g. see MACRO application for logistic regression in PROC NLIN, SAS Institute Inc. 1989). However, any of these *a priori* transformations require that the nature of the relationship is well understood and able to be approximated, because it is not appropriate to just include them as options for selection in modelling. Certain knowledge of such relationships for whole-plant response is rare in ecological modelling, but is the objective of physiological process models. In disciplines, such as predictive microbiology, the relative ease of multifactorial experimentation offers some clues as to the underlying non-linear form of the response to particular environmental gradients (e.g. Ratkowsky & Ross 1995).

In the absence of more precise information, polynomials offer a simple and flexible means of approximating the non-linear shape of species' or genotype responses to an environmental gradient. For example, Scheiner (1993) suggested the use of polynomial terms as a means of approximating a reaction norm function of a genome relative to an environmental gradient. However, Prentice *et al.* (1991) expressed reservations about the use of polynomials for defining the response surfaces of species' migration patterns in North America. They commented that polynomials tended to fit trends which were inherent to the data rather than necessarily indicative of plant responses. Sampling bias exacerbated this problem in their analysis, leading to better fits to the data in some regions of climate space than in others. In addition, they found that polynomials limited the potential for extrapolation, and chose to use the LOWESS method instead (after Cleveland & Devlin 1988).

Polynomial regression is a form of trend surface analysis. It is a means of deriving a response surface for species' distributions by interpolating existing patterns. For example, Mucina *et al.* (1991) discuss polynomial regression as a tool for depicting quantitative vegetation and ecological data on maps. Similar techniques have been used to map other landscape attributes of soils (e.g. McKenzie & Austin 1993; Moore *et al.* 1993a) and climate (e.g. Guenni *et al.* 1996; Guenni 1997). In the context of trend surface analysis, Mucina *et al.* (1991) suggest that polynomial regression is a useful way of transforming irregularly patterned data by interpolation onto a regular grid, also termed regularisation. While their method of defining algorithms for polynomial splines arises from the simple purpose of mapping biological data, it is analogous to the use of polynomial transformations of environmental gradients for modelling species' distributions. In this latter case, trends in response, derived by regression modelling of the irregular samples of species' distributions with polynomials of component environmental gradients, can be viewed by interpolating onto a regular map grid of the same component gradients. The final map of predicted distribution patterns represents a regularisation of the original sample data, but has the advantage that it can be examined for ecological inference in the context of the component gradients.

Therefore, until other forms can be shown to be functionally consistent, polynomial terms for explanatory variables seem to be the most appropriate means of approximating species' responses in regression models. Since complex shapes of bimodal and highly skewed responses are experimentally and theoretically possible (e.g. Fresco 1982, Austin *et al.* 1985), polynomials up to the fourth order need to be considered. However, some complex response shapes may be artifacts of data inconsistencies, such as sampling bias or an incomplete description of the response due to the omission of key explanatory variables (Austin *et al.* 1990). A representative sample, or explicit recognition of the nature of sampling bias and its implications for correlative model fitting, is needed because polynomials tend to fit trends that are inherent to the data, rather than those which are necessarily indicators of actual plant response. Some exploratory data analyses may help assess which responses shapes can be realistically considered in parametric models (e.g. Yee & Mitchell 1991; Leathwick 1995).

1.8.2.6 Specifying interaction terms

Plant responses to environmental gradients are known to be factorial, although experimental assessments are frequently limited to simultaneous testing of only two- or three-way interactions (e.g. Hawkins & McDonald 1994; Henderson-Sellers & Henderson-Sellers 1996; Lippert *et al.* 1996; Walters & Reich 1996). For example, an interaction between soil nutrient status and soil water status is expected to influence the floristic composition and productivity of Australian eucalypt forest communities (Florence 1996, p. 38). Interactions between two or more environmental gradients are therefore expected to be determinants in species' distribution patterns (e.g. Austin *et al.* 1990, 1996; Leathwick & Mitchell 1992; Cloutman & Cloutman 1994). However, Austin *et al.* (1990) found that their interaction terms were rarely significant in

factor models, and Major (1951) suggests that 'the great variety of combinations in nature renders the effects of one factor independent of the other'. In any case, there seems little *a priori* reason to choose a particular interaction, or set of interaction terms, given the necessarily large range of explanatory variables needed for modelling species' distributions (although see application of numerous interaction terms by Lenihan 1993).

Statistically, parsimony provides the criterion for selection of terms in model building. Two-way interactions might be considered before three- or four-way interactions. With relatively few explanatory variables (e.g. up to three or four), the range of possible interactions terms could be considered as candidate variables in ecological modelling. However, if the functional relationship between the response and explanatory variable is defined by polynomial terms, then this complicates the specification of interaction terms. A large number of candidate variables further complicates possibilities for considering combinations of interactions in ecological models (McCarthy *et al.* 1996).

Environmental gradients, such as rainfall, evaporation, temperature and solar radiation, which are meteorological estimates of the climate in a region, are known to be codependent, but the nature of this dependency is location-specific at a regional scale (e.g. Neilson *et al.* 1992). For example, in coastal regions, temperature may vary to a minor extent with changes in rainfall because of the ameliorating influence of ocean currents. However, in inland regions, where the climate is not so buffered, temperature and rainfall may be more closely correlated. Therefore, models of species' distributions as a function of climate may only be applicable to the regions from which the sample were derived, unless interaction terms are also included. Interaction terms may be able to account for regional differences in the relationships between climate variables, such as temperature and rainfall, or rainfall and soil genesis (e.g. Grant *et al.* 1995), and thus be applicable to the extrapolation of predictions (e.g. Leathwick & Mitchell 1992; Austin *et al.* 1996).

However, it is not yet computationally feasible to consider the range of possible interaction terms that may be significant in defining the distribution of a species. There is no *a priori* reason for selecting one form over another. Current ecological modelling therefore largely avoids the use of interaction terms, or only considers a few possibilities (e.g. Austin *et al.* 1996). As a result, predictions are restricted to interpolations within the range of the explanatory variables included in the model.

1.8.2.7 Multicollinearity between explanatory variables

The relevant variables to be used in regression analysis are not necessarily known prior to analysis and an array of candidate variables are typically tested for significance. These variables are often highly correlated, leading to problems of multicollinearity — a near-linear relationship among the explanatory variables. The effect of multicollinearity is manifested by extraordinarily large estimated standard errors and sometimes by a large estimated coefficient as well (Hosmer

& Lemeshow 1989), restricting the applicability of the model. Multicollinearity in a data set may arise due to physical constraints on the model, or in the sampled population, biased sampling, or an over-defined model (e.g. MacNally 1996). In multivariate relationships, natural variability in one variable can induce statistical significance in collinear variables even if they bear no structural relationship to the response (Prairie *et al.* 1995).

Backward and forward selection are standard regression approaches for dealing with the selection of a subset of independent variables from a highly correlated set of candidate variables. Partial correlation is also commonly used for assigning likely influence of independent variables. However, there is a trade-off between the explanatory power of models and their complexity — statistics which help define logistic regression efficiency include the Akaike Information Criterion and the Schwartz Criterion (e.g. SAS Institute Inc. 1997). Other statistical approaches, such as Hierarchical Partitioning (Chevan & Sutherland 1991; Christenson 1992), which estimate the explanatory capacities of a set of independent variables, provide a supplementary interpretive tool in multivariate inference (MacNally 1996, 1997; see also Candy 1997).

1.8.2.8 Statistical model building

From the previous discussion, it is clear that statistical model building requires decisions about the shapes of responses, the potential for significant interactions between explanatory variables and the method for selecting the set of explanatory variables that best describes the ecological response. Forward stepwise selection methods have been suggested as a systematic approach to model building (e.g. Nicholls 1989, 1991; Austin *et al.* 1990; Austin & Meyers 1996). However, backward elimination methods, which take into account the simultaneous relationships between variables and their potential shapes, may be more appropriate where a subset of candidate variables represents the potential for defining a maximal model (e.g. Leathwick 1995).

The fact that polynomial or interaction terms might be included in a multiple regression model also influences the way in which variables are selected and the criteria which might be used in deciding between competing models (McCullagh & Nelder 1989; Nelder 1994, 1995, 1996). For example, in the case of polynomial terms, Nelder emphasises that fitting a higher order polynomial, such as a quadratic, without also fitting the lower order term (e.g. linear function) is usually unhelpful. This is because the absence of the linear term implies that the maximum (or minimum) of the response occurs at zero values of the explanatory variable. Ordinarily, there is no reason to suppose that the turning point of the response is at a specified level of the explanatory variable (McCullagh & Nelder 1989). The same argument applies to interaction terms between covariates: interaction terms should always be included with their corresponding linear terms because there is no reason to postulate special properties for the origin (McCullagh & Nelder 1989). Such terms should always be entered into or removed from a model simultaneously.

Nelder (in McCullagh & Nelder 1989) indicates that the relationships among polynomial terms for a covariate are similar to those among factors and interactions. Their functional marginality is not a true marginality, because no linear dependencies among covariates are involved, but their functional marginality does impose constraints on the order in which terms should be introduced into a model. Therefore, some existing automated methods of variable selection for model building may not be appropriate. Selection criteria that are consistent with ecological theory and current statistical concepts may need to be manually applied. For example, when including or removing polynomials during stepwise model building procedures, selection criteria should always be applied to the significance of the highest order polynomial term for a covariate. Similarly, the significance of a two-way interaction term would need to be assessed in the presence of its linear terms. Non-significant values of the linear terms (or the lower-order polynomial terms) would need to be ignored in favour of the significance of the interaction term (or the higher order polynomial term). Where these lower order terms are included for marginality reasons but reveal very poor fits, other problems of collinearity may exist in the model. Tests for improving model fit or parsimony by removing the higher order term, even though apparently highly significant, could be tried.

Since polynomials represent approximations to an underlying shape, it seems sensible to adhere to Nelder's simple criteria which guide a controlled and consistent approach to variable selection and model building (although see discussion and rejoinders by Rodriguez *et al.* 1995; Searle 1995; Aitkin 1995; Lindsey 1995; Gower 1995; van Eeuwijk 1995). However, it should be noted that the data are unlikely to indicate one solution. Clustered around the 'best' model will be a set of alternatives, almost as good and not statistically distinguishable (McCullagh & Nelder 1989).

1.8.2.9 GLMs versus GAMs

Generalised additive models (GAMs) have become popular in recent times as an alternative to generalised linear models (GLMs) because they provide greater flexibility in the choice of transformation of explanatory variables (e.g. Chambers & Hastie 1992). GAMs are an extension of GLMs, allowing a wider range of parametric, semi-parametric and non-parametric terms to be defined in the model structure. In this respect, GLMs become a special case of GAMs in which all terms are defined by parametric methods.

Yee & Mitchell (1991) introduced GAMs to plant ecology as an exploratory tool in the analysis of species' distributions with respect to climate. They suggest that because GAMs are data-driven rather than model-driven, GAMs provide opportunities for examining the structure of the data as a means of deciding which transformations of explanatory variables might be appropriate for testing in parametric modelling. In this context, GAMs provide a means of hypothesis testing (Yee & Mitchell 1991). Scatterplot smoothers of various types can be applied (e.g. Hastie & Tibshirani 1990) to reveal data features (e.g. bimodality and asymmetry) which may have ecological interpretations. However, the desire to reveal the structure of the data for the purpose

of ecological inference assumes that the given sample is completely representative, otherwise any functional interpretation of shape may be spurious.

Norton and Mitchell (1993) discuss GAMs in the context of wildlife habitat modelling and population viability analysis. They suggest that that GLMs may not be appropriate for analysing field observations of biological pattern because many responses of wildlife to environmental factors are non-parametric in form (see also Norton & Williams 1992). Their criticism of GLMs is largely derived from the belief that parametric methods, such as logistic regression, are unable to assess whether the response curve of a target species to environmental factors, in one dimension, is symmetric and bell-shaped or not. This is not a valid presumption for deciding between the use of GLMs or GAMs.

Further studies on the relative merits of GLMs and GAMs for species' distribution modelling were conducted by Austin and co-workers (Austin *et al.* 1994; Austin & Meyers 1995, 1996), and by Ferrier and co-workers (Ferrier & Watson 1996; Ferrier & Pearce 1996). Austin & Meyers (1995, 1996) conclude that both GLMs and GAMs can produce useful predictions where the choice of predictors is based on direct environmental gradients, and attention is given to the appropriate definition of a sampling domain. They found that restricting the range of the data played a significant role in improving predictions, because of the distorting influence of 'naughty-noughts' on response functions in either case. Although the relative gains to spatial modelling offered by either technique appear to be equivocal, Austin & Meyers (1995, 1996) suggest a mixed strategy of GLM and GAM methods, depending on which appears to perform the best for a given set of data and environmental predictors.

Ferrier and co-workers, however, intially found that GAM models performed better than GLM models (Ferrier & Watson 1996) for a range of fauna and vegetation response variables, but in later work (Ferrier & Pearce 1996) concluded that there was no significant difference in the predictive performance of either technique. Rather, they found that data characteristics were more likely to influence the performance of a predictive distribution model. For example, factors of sample size, species' rarity, and the presence of geographic or environmental bias could reduce the reliability or discriminatory performance of a model (Ferrier & Pearce 1996). They conclude that regional species' distribution modelling (with either GLMs or GAMs) can provide a reliable approach to estimating distribution patterns for regional conservation planning, but that validation with independent data of similar quality and geographic extent is needed to quantify predictive reliability.

Since data characteristics appear to be more important than the modelling technique when deriving a predictive model, and GAMs are more likely to fit spurious response functions to biased data, I used the parametric approach of GLM for modelling species' distribution patterns.

1.9 Presenting the modelling results

For models correlating species' distribution patterns with environment, the estimated probability of occurrence represents a statistically derived performance index. However, the lack of informative presentations of the results of such models have limited their usefulness for interpretation by land managers and researchers (Norton & Williams 1992). Model predictions represent a site index for the occurrence of the species, and as such are commonly projected onto a geographic map. Another way of viewing the predicted distributions is to display them as a set of responses along an environmental gradient. The latter form of presentation is consistent with the methods of direct gradient analysis developed by Whittaker and others (e.g. Whittaker 1956, 1967, 1975; Austin 1980, 1986, 1987) as a means of interpreting the individualistic responses of species.

In either case, the characteristic responses can only be realistically viewed in the context of the availability of habitats in the landscape. This is because the nature of a species' response, as interpreted from its distribution patterns, depends upon the availability of particular combinations of environmental factors; not all of which may exist in the landscape and some of which are constrained by the inter-dependence between climatic factors. The modelled response is based upon a sample of these environments, but needs to be interpolated onto a regular geographic grid to remove the potential for further extrapolation of sampling bias. The levels of each factor in the model therefore need to be known for each point on the interpolation grid.

1.9.1 Direct gradient analysis

The ecological interpretation of a statistical model of species' distributions requires methods of display that are consistent with the continuum concept. The continuum concept considers species' responses as a function of an environmental gradient (Austin & Smith 1989). Early interpretation of the shape of plant species' responses relative to environmental gradients was based on simple free-hand descriptions (e.g. Whittaker 1956; see review in Westman & Peet 1982). Logistic regression analysis has subsequently provided a statistical tool for flexibly and objectively describing these response shapes (e.g. Austin *et al.* 1984; Bartlein *et al.* 1986; Yee & Mitchell 1991). However, Norton and Williams (1992) suggest that these logistic regression models are limited in their ability to test whether the response curve for a species, in one dimension, is symmetric, bell-shaped or some other form (see also Norton & Mitchell 1993).

The characteristic manifold response of a species' realised niche from a logistic regression model can be displayed in one dimension. The single gradient ecological responses consist of an upper bound to physiological adjustment, or constraints due to competition and other ecological or biogeographic processes, and a continuum in possibilities reflecting declining plant performance caused by limiting levels of other environmental factors. Methods of graphic display that dissect this scatter diagram to reveal the internal structure are needed if the results of these predictive models are also to allow ecological inference.

The problem of interpreting the nature of ecological responses within a cloud of possibilities was also considered by Thomson *et al.* (1996). They recognised that multiple factors associated with species' distributions lead to univariate or bivariate plots with a scattering of data points beneath an upper limit. Because of its interpretive significance to the ecology of the species, a direct description of these responses was needed. They suggested that previous workers had erroneously dealt with these patterns as 'triangular distributions' (e.g. Maller *et al.* 1983). To distinguish their analysis, they coined the term 'Factor Ceiling', or ecological ceiling, to describe this characteristic response. Their observations were derived from an experimental study of the spatial patchiness of flowering and seedling performance of a lily in subalpine meadows.

Thomson *et al.* (1996) found that conventional methods of correlative analysis, such as regression, could not characterise the upper limits of a species' response. They demonstrated methods of partitioned regression and logistic slicing as a means of examining the form of a response in a factor ceiling cloud of responses. Similar solutions to the problem of analysis and interpretation of such data were developed by Blackburn *et al.* (1992). Blackburn *et al.* (1992) used the highest response value in each vertical 'slice' of a factor gradient in body-size classes as an indication of the position of the outer envelope of the abundance response of beetles. They then fitted a regression line through those points to estimate the outer boundary of the response. A statistical estimate for the form of the upper limit to response along a gradient could thus be objectively defined, although this was not regarded as a significance test of the response shapes (Thomson *et al.* 1996).

For the purpose of presenting the results of a predictive model for the relationship between species' distribution patterns and environment, I adopted graphic display techniques of direct gradient analysis. I developed simple descriptive methods to reveal how the model defines species' responses with variation in the environment. Particular percentile responses in each of 10 to 15 vertical slices of the cloud of responses along an environmental gradient were linked by a spline curve to reveal the characteristic form of the modelled response; as a representation for the continuous series of nested curves. The outer envelope reflects the maximum ecological response with respect to the environmental gradient when all other combinations of factors are non-limiting. The sensitivity of the response to variously limiting conditions of each environmental gradient, akin to simulating an experimental design, can thus be explored. In this respect different hypotheses of the relationship between a species' occurrence and its habitat can be considered, taking into account the fact that possibilities for variation between the different gradients of climate and substrate are not infinite and can be defined.

These analyses differ from those pursued by Thomson *et al.* (1996) and Blackburn *et al.* (1992) in that the underlying features which lead to the response are known from the statistical model. The objective of the direct gradient analysis is therefore to reveal the shapes of the response for comparison between species, or other ecological grouping, and between physiological and ecological responses. The characteristic shapes can be subsequently considered in the context of

ecological theory (e.g. Austin & Smith 1989; Austin & Gaywood 1994). Clearly, parametric methods of generalised linear modelling can be used to explicitly test fundamental aspects of ecological theory, such as the characteristic form of a species' response to environmental gradients in one dimension (cf. Norton & Williams 1992; Norton & Mitchell 1993).

1.9.2 Mapping predictions

Interpolation of the results of the statistical model as a geographic map of the predicted distribution is a simple way of presenting complex patterns (e.g. Bartlein *et al.* 1986; Lenihan 1993; Carey *et al.* 1995a; Austin & Meyers 1996). However, while this provides a visual evaluation of the model by comparison with known occurrences, it does not facilitate ecological interpretation of the results. Nevertheless, mapped predictions can reveal weaknesses in the underlying model due to sampling bias or missing variables leading to inappropriate polynomial shapes for some gradients. For example, spurious positive responses may reveal limits to interpolation where inadequate sampling for particular environmental combinations results in effective extrapolation in some geographic regions. Therefore, the appropriate interpolative domains for prediction will be determined not only by the method of statistical fitting (Prentice *et al.* 1991), but also by the regions of environmental space in which representative sampling was achieved. Spurious extrapolation into environmental space that was poorly sampled, and unrepresentative of its landscape occurrence, will also influence the patterns displayed by, and subsequent interpretation of, the direct gradient analysis.

Maps of species' responses derived by standard statistical inference can also be used to demonstrate the confidence limits associated with the predicted distributions (Ferrier & Watson 1994). For example, three maps could be generated to depict the estimated probability of occurrence and the upper and lower 95% confidence limits (or other confidence limits). However, Ferrier & Watson (1994) caution against the interpretation of these confidence limits because they measure prediction error by simple resubstitution, rather than with an independent sample, and therefore potentially under-estimate the true prediction error of the model. Nevertheless, the presentation of associated maps for comparison of prediction error gives an immediate perspective on model validity and robustness, a simple and important interpretive tool for land managers and ecologists (e.g. see discussion by Norton & Williams 1992).

1.10 Interpreting species' responses

The interpretation of species ecological responses from a parametric model makes assumption about the validity of statistical and ecological premises on which the predictions are based. Inconsistencies between statistical assumptions and ecological theory may confound the interpretation of the shapes of species' responses to environmental gradients, or require a number of different approaches to confirm inferences. For example, ecological theory implies that competition occurs between species, and therefore species' responses will also be correlated. The

statistical models also assume that the sample is representative and that the explanatory variables account for all systematic types of variation. The correlative analysis further assumes that the scale of observation for the response and explanatory variables are the same. These may not be the case, but a review of the possible deviations could help explain anomalous modelled response patterns.

1.10.1 Explicitly accounting for competition between species

Current approaches to the analysis of species' distributions based upon regression do not take into account the correlation between species. Models deal with one response at a time. Nonetheless, species' responses may be indirectly correlated because of a coincidence in the relationship with particular environmental factors, or they may be directly correlated due to competitive interactions. Other applications of regression analysis are needed to simultaneously and explicitly account for the correlations between species and their relationships to environment (M. P. Austin personal communication).

Various approaches to this problem of statistical analysis of correlated responses have been developed from simultaneous equations theory of econometrics (e.g. Brzeziecki 1987), structural equations (e.g. Grace & Pugesek 1997), or by iterative solutions through canonical correspondence analysis (e.g. ter Braak 1986). New directions based on variance partitioning (e.g. Breiman & Freidman 1997), or multivariate statistical techniques and non-parametric regression (e.g. ter Braak & Juggins 1993; ter Braak 1995; ter Braak & Verdonschot 1995; Doledec *et al.* 1996), are evolving. However, current approaches to spatial prediction modelling remain largely a matter of subjective comparison of species' response shapes. It is also possible to examine levels of correlation between responses. However, in the absence of the necessary information for constructing a structural equations model, iterative analyses of species' distribution patterns could be developed, as an interim method for explicitly accounting for the correlation between species' responses.

An iterative analysis of the putative positive or negative codependency for the distribution of two or more species with similar geographic distributions is theoretically possible, but is expected to be costly in both the physical and computational times required to develop such a model. The iterative technique might involve constructing separate models for the initial relationship between two or more species' responses and features of their environment. The predicted responses in each case could be applied as dependent variables for each of the other species, the models re-built, and the results again applied back to each; now up to a third iteration phase. Iterations would continue until no significant improvements in model fit occurred with the inclusion of the predicted responses from any of the other species as dependent variables for the distribution of the target species. In this way, the influence of another species associated with the target species may be determined as a function of the same set of environmental predictors. However, the correlative nature of the analysis means that the final response function does not

necessarily separate neutral effects from real interactions, especially where there is a large degree of coincidence in the environments occupied by either species. One may only assume that species with similar ecological ranges are likely to be competing if processes are operating at the same scale, or may be neutral or facilitative in their interactions if they partition the environment at finer scales than the modelled gradients.

1.10.2 Considering differences in scale between response and explanatory variables

A number of factors confound the interpretation of an ecological response. In particular, ecological observations are frequently mismatched between the scale at which the response is observed and the scale at which the environment is recorded (Palmer & Dixon 1990). The sample of the response which records the presence or absence of a plant within a plot is generally heterogeneous, but the observation of the environment assumes within-sample homogeneity. The degree of heterogeneity of a plot is unknown, but will vary between plots. This variation may be systematic when considered in the context of the experience of environment by an individual species (Neilson *et al.* 1992).

This mismatch of observation scales between the plant response and the environment may be exacerbated by differences in the accuracy and scale of estimates for gradients in climate and substrate type. For example, the climate of a vegetation sample (e.g. 0.1 to 0.3 ha) is usually approximated from physical process models which generally estimate variables at a much coarser scale (e.g. 200 m grid - McMahon *et al.* 1996). The climate estimation procedures may also involve a degree of systematic error which may exacerbate the mismatch between plant response and its proximal environment. Errors associated with weather estimates vary with the complexity of topography and the density of the meteorological network in the target region, but are normally < 0.5C for monthly mean temperature and < 10 % for mean monthly precipitation (Hutchinson 1987, 1989). This limiting resolution for the environmental gradients may therefore contribute to the level of predictive error.

The difference between the scales at which the plant response and its environment are observed may have several consequences for the ecological interpretation of subsequent models. Scale in this context reflects spatial attributes of habitat heterogeneity. For example, Neilson *et al.* (1992) develop a theory of hierarchical constraints due to climate, substrate and biotic interactions on the spatio-temporal patterning of habitats and therefore their constituent species:

> *Climate remains homogeneous over a larger spatial extent than does the substrate, which is usually homogeneous over a larger spatial extent than individual organisms. Local variations in topography and substrate produce a variegated pattern of microclimates that modulate local plant distributions. For example, two ecologically similar species may exhibit considerable overlap in their distributions at the biome or regional scale, but at the local scale, the two species' distributions may overlap very little as competitive interactions could eliminate one or the other on any given substrate, or under particular microclimates. Clearly, different processes are operating to produce overlap at one scale and segregation at another. An envelope of constraints will define the region within which two species may coexist. However, the physical mapping of those constraints on a variable landscape will determine the spatial scales of suitable habitats at different places within the region. (p. 145)*

In the context of interpreting species' ecological responses from a direct gradient analysis of the predictions, take the case of a long tailed response of a species to an environmental gradient. This form of response may indicate that the species has a wide ecological tolerance due to the persistence of populations in currently unsuitable types of habitat. For example, long-lived species (e.g. *Athrotaxis* species, Cullen & Kirkpatrick 1988a, b; Cullen 1991) could maintain occupancy of a site even though regeneration is unlikely to be successful, being relicts of past types of climate, or examples of biological inertia (*sensu* Neilson *et al.* 1992). Alternatively, the tailed response may indicate the existence of sparse occurrences of micro-habitat which are closer to the types of environment represented by the optimal response of the species. These micro-habitats may be quite poorly correlated with the mean record of environment for the site used in modelling distribution patterns. Therefore, the long-tail of an ecological response may be an artifact of the degree of difference between the proximal environment experienced by the population and the average environment recorded for the plot.

At the margins of a species' distribution, the species' experience of environment may be one of increasing heterogeneity, even though for another species with a different physiological response, this environment will be experienced as relatively homogeneous (e.g. see discussion by Neilson *et al.* 1992). Recognition of the possibility of these effects simplifies the interpretation of empirical models of species ecological responses. For example, in the absence of known micro-site variation, genetic drift, and mass effects (e.g. Schmida & Wilson 1985; van der Maarel 1995; Jobaggy *et al.* 1996), the long-tail of an ecological response may be theoretically interpreted as reflecting competition intensity, or physiological stress (e.g. Austin 1990; Austin & Gaywood 1994). As a result, prediction of marginal distributions from the tail of an ecological response may have much less accuracy than predictions in the region of the optimum response. For this reason, Lenihan (1993) chose to develop models based upon the probability of species' dominance, rather than occurrence. He suggested that species' dominance was more appropriately scaled to the regional macro-climate predictors that were available for modelling. It was thus expected that the resulting models would provide a description of the climatic parameters associated with the physiological tolerance of the species.

Neilson *et al.* (1992) also discussed the possible types of competition that might be expected under different habitat regimes. They suggest that in core regions of a physiognomic vegetation type areas of optimal microhabitat are spatially contiguous and extensive, so diversity of related

species might be relatively high due to niche packing, but competition would be diffuse and competitive exclusion less likely than in marginal habitat. Conversely, in the marginal habitats the spatial extent of suitable micro-habitats would be small and patchy and ecologically similar species might face more direct competition, and in some cases, competitive exclusion. In this latter case, the stringent environment constrains the coexistence of ecologically similar species to fewer species.

These types of interactions have implications for modelling species' distributions and interpreting ecological response patterns. For example, a simple model for the probability of occurrence of a species may be inconclusive. But successive models that develop different ecological themes to demonstrate a consistency in the relationship between response patterns and environment may be more conclusive. Separate consideration of species' occurrence or dominance, and the manner of co-occurrence with other species may clarify the functional parameters of a species' distribution.

1.10.3 Some ecological considerations for the interpretation of response shapes

The overall shape of the response has implications for the potential match between ecological and physiological optima, and the direction of skew has implications for interpreting the relative importance of biotic and abiotic factors (Austin 1990; Austin & Gaywood 1994). A Gaussian shape of the ecological response to an environmental gradient may be indicative of biotic effects that constrain the species to a central position of its physiological response. This assumes that the physiological response is also Gaussian and broader in width than the ecological response. A skewed physiological response has different implications for the interpretation of the Gaussian ecological response.

The symmetry of competition with other species also has implications for the shape of the ecological response relative to its physiological potential (Austin 1990). Assuming the potential response is also Gaussian and competition is symmetric, the optimum of the Gaussian ecological response is likely to be a close approximation to the physiological response. Assuming competition to be the dominant factor that constrains the species from occupying positions of the environmental gradient close to its physiological limits, then different species are involved in competition on either side of its optimal ecological response. The species with a Gaussian response may be seen as an effective competitor, through resource exploitation, in the region of the environmental gradient which approximates its physiological response.

The ecological interpretation of skewness also requires comparison with the shape of the underlying physiological response, to assess the degree of deviation. If the overall ecological response is skewed with respect to the physiological response, then competition is expected to be responsible for excluding the species from achieving its physiological potential in the region of the tail (e.g. Fresco 1982). With skewed shapes, the optimum ecological response is also expected to be shifted from its physiological optimum. Shifts in optima are expected to reflect

competitive ability associated with the tolerance of particular environmental conditions, rather than competitive ability based on resource exploitation.

The degree of skewness, while potentially indicative of scale or data problems, may also have implications for the intensity of competition. In such cases, more than one species may be involved in competition, resulting in a hierarchy of responses along the environmental gradient which increases the distance between the physiological and ecological optima. However, while a deviation between ecological and physiological optima is theoretically more likely in the case of skewed ecological response shapes, the degree of deviation cannot be inferred without reference to experimental studies which define a shape for the plant physiological response. Some knowledge of the shape of the underlying physiological response is needed to clarify the interpretation of ecological response patterns.

1.10.4 Inferring site abundance from predicted logistic performance

If the estimated probabilities of occurrence from logistic regression models could be interpreted in terms of, say, the relative abundance of a species at a site, then the predictions from an analysis of species' distribution patterns have widespread application to questions of land management (Ferrier & Pearce 1996). For example, prediction of species density could be used in reserve design to optimise selection of areas with respect to the likelihood of reproductive success, critical refugia, or maximum biodiversity attributes. Alternatively, predictions could be used to assess sites for optimum growth of commercial species, as distinct from areas where survival is ensured but productivity is low. In this respect, the predicted probabilities of occurrence could contribute to a population viability analysis (e.g. Norton & Mitchell 1993). Thus interpolative predictions of the relative occurrence of plants could be used as management tools for assessing conservation values or options for land use, and the needs for further planning or more narrowly focused research.

The potential for interpreting species' density from predictive models of distribution patterns was investigated by Ferrier & Pearce (1996). They tested the degree of correlation between observations of species' relative abundance and predicted probabilities of occurrence derived from a presence/absence model using a Spearman rank correlation coefficient. They found a positive correlation in many cases, particularly in comparison with observations of absence/relative abundance, but presumed this to be a consequence of simply a presence/absence correlation. They preferred to base their conclusions on comparisons of predicted probabilities with presence-only observations of relative abundance. For this case, they conclude that models with a high discrimination capacity had a greater likelihood of describing relative abundance, but that in general predictive models did not readily identify habitat with a high relative abundance of the species in question. They suggest that functions of species' density are more closely associated with habitat characteristics at a finer scale of resolution (e.g. competitive and social interactions between species) than those used to model species' distributions.

However, it may not be statistically valid to compare a model based on the logistic function of presence/absence data with an observed relationship based on the relative performance of a presence response. The Spearman rank correlation coefficients for the comparison of predicted probabilities with observations of absence/relative abundance is more likely to give a valid analysis of the question of the suitability of interpreting predictive models in terms of species' density. In this case, there is a consistent trend for the comparisons presented by Ferrier and Pearce (1996). Further analyses and comparisons are needed to clarify the interpretability of spatial prediction models for species' distributions in terms of the finer-scale population responses at a site. At the very least, positive trends in the comparison of such relationships suggest the use of predictive models as guides for the selection of sites and the design of experiments to clarify the abundance and demographic behaviours of species. The question of the abundance interpretability of species' distribution models is therefore considered in this thesis.

1.11 Conclusions

This review has revealed how information about species' physiological and ecological responses from experimental results and numerous observational studies can be used to interpret species' distribution patterns from a correlative analysis of their occurrences and the environment. Since correlative methods essentially aim to fit the relationships inherent in the data, it is important that the sample of species' distributions (or other performance measure) be completely representative of the relationships being studied. If this is not possible, then the bias inherent to the sample should be examined and explicitly stated prior to analysis, so that subsequent interpretation of response patterns can be considered in this context.

Predictive models thus allow plant distribution patterns to be interpreted as differential responses to environmental gradients which influence the over- and under-supply of essential resources and conditions for plant growth, regeneration and survival. The presentation of predictions as interpolated response surfaces (maps or direct gradient analysis) complements the explanation for species' range limits and relative performance derived from physiological process models.

Empirical interpolation of plant distribution data is also a practical tool for land management. Land managers are interested in predicting the presence, absence or abundance of a species at a given site, or the general trends of distribution throughout a management zone. They may also wish to assess the likely outcome of an altered management regime upon the temporal and spatial scales of vegetation pattern.

Ecologists, however, are interested in understanding — testing theoretical concepts and developing new ideas about how plants interact with each other and their environments, and then using this information to refine factorial experimental designs. The shape of the ecological response from empirical models of species' distributions has implications for interpreting the potential mechanisms associated with either biotic or abiotic influences. Theories of niche

differentiation and reaction norms also suggest that species will tend to evolve their responses toward the norms of their environment. The overall complexity of shape (skewness and multimodality) and width of a species' optimum ecological response (plateau or peaked shapes) for each environmental gradient may further indicate the nature of genetic and phenotypic traits, or patterns of biological interference, that underlie each type of response.

1.12 Thesis outline: A systematic approach to predicting plant distributions

Within this thesis, I develop a systematic approach to predicting plant species' distribution patterns. Inherent to this approach is the consideration of practical, statistical and ecological constraints on the structure of the sample and approaches to analysis. Though most importantly, I develop methods of display and presentation to facilitate an interpretation of the results which links models of species' distribution patterns with known physiological processes and current theoretical concepts in ecology (Box 1.1). These methods are developed around a set of *ad hoc* compiled data for the distribution of *Eucalyptus* species in Tasmania using the standard statistical procedure of logistic regression, representing an application of generalised linear modelling.

A set of data for the occurrence of *Eucalyptus* species in Tasmania was compiled from a number of different sources, where consistent levels of information needed for prediction were available. Since these data represent an *ad hoc* collation, an exploratory analysis of their sampling adequacy was undertaken. Initially, the extant areas of eucalypt forest habitat in Tasmania were compiled to provide a context for assessing their ecological variability. The necessary levels of sampling for eucalypt occurrences that might be expected from these habitats were derived by applying the logic of a regional biophysical survey design (Chapter 2). However, while this analysis provides a perspective on the regional sampling adequacy of eucalypt habitats, it does not define the sampling domains for analysis of individual species' distributions.

An objective definition of species' sampling domains was needed to avoid the anticipated analytical problem of 'naughty-noughts' (e.g. Austin *et al.* 1996). For this purpose, the potential geographic and environmental ranges of Tasmanian eucalypts were clarified by collating all available sources of species' presences that could be validly located within 100 km^2 grid cells (Williams & Potts 1996 — attached with this thesis). These distribution ranges were then used to distinguish the sampling domains for each *Eucalyptus* species within the compiled set of ecological data (Chapter 3). In this context, the atlas of species' distributions filled a second role — providing a basis for assessing the sampling adequacy of individual species' occurrences. Exploratory data analysis methods were used to assess sampling ranges and evaluate local representativeness. These analyses provided insight into the levels of sampling bias associated with the presence and absence records for each species. The results have application to the interpretation of ecological responses from subsequent models based on these data.

To maximise the ecological relevance of predictive models, gradient analyses require that the environmental scalars be as proximal to plant physiological processes as possible. Environmental gradients in water, temperature, nutrients and light directly influence plant physiological processes. However, the existing set of explanatory variables, such as rainfall, evaporation, mean annual temperature and substrate type, generally reflect average, ambient conditions, rather than proximal gradients in resource supply. The potential for a more proximal gradient in soil water supply to improve the explanatory power of ecological models over less direct gradients in climatic variability was assessed (Chapter 4). The application of such a water supply gradient for comparative database and experimental studies was subsequently demonstrated for an ecotone between two related *Eucalyptus* species: *E. tenuiramis* and *E. obliqua* (Chapter 5).

The potential for a similar process-orientation of temperature and solar radiation variables in the context of the plant physiological response, assayed as the average rate of photosynthesis and carbon gain for *Eucalytpus* species, is also assessed for improvements in prediction and ecological explanation (Chapter 6). The objective of this analysis was to determine whether productivity gradients defined from known physiological processes could replace component direct environmental gradients in predictive models. Improved model fits would indicate that the multivariate character of the environment could be collapsed into fewer gradients, thus clarifying the ecological response of a species and facilitating comparisons of the realised niche for different species.

The results of these analyses are applied to a description of the realised niche of *E. globulus* in eastern regions of Tasmania (Chapter 7). Initially, the sampling adequacy of the data for modelling the occurrence of *E. globulus* is reviewed. Patterns of co-occurrence and dominance between *E. globulus* and other species within its geographic range are summarised to assess the potential for competitive, neutral or facilitatory interactions. The univariate ecological response of *E. globulus* to major biotic and abiotic gradients in its environment are explored. Finally, predictive models that combine the effects of different environmental gradients in water, temperature, nutrients and light, are correlated with the distribution of *E. globulus*.

Models for the distribution of *E. globulus* occurrence and relative dominance were compared, as were the improvements in model performance following the inclusion of biotic attributes with abiotic predictors (Chapter 7). These comparisons of predicted occurrence and dominance provide a means of assessing the relative potential for interpreting species' density or abundance at a site from estimated probabilities of occurrence based on simple presence/absence data. Furthermore, comparisons of the relative merit of biotic and abiotic predictors are expected to indicate the relative importance of undefined environmental factors associated with the species' habitat. These effects might be due to scale differences in the experience of the micro-habitat, or species' interactions, or historical biogeographic and landscape processes, including disturbance regimes.

The potential for different ecological responses to be associated with broad variation in vegetation type were also examined (Chapter 7). Significant differences in the ecological response of *E. globulus* in forest stand types classified as wet sclerophyll or dry sclerophyll might indicate underlying genetic differences in the adaptive traits being expressed by the phenotypes as an overall bimodal distribution response. Alternatively, a bimodal response shape might be a consequence of competition with an ecologically similar species. The likely effect of competition by related white gum species (*E. viminalis, E. dalrympleana* and *E. rubida*) was subsequently assessed by considering differences in performance of *E. globulus* and the white gums in different stand types defined by the presence, absence or co-occurrence of either species. These analyses enabled hypotheses of forest ecology relating to the natural distribution of *E. globulus* to be explored, including the potential for competitive exclusion and the relative influence of chance and biogeographic factors in determining distribution patterns, super-imposed upon climate and substrate effects.

In deriving these realised niche models, the rigorous evaluation methods of Ferrier and co-workers was followed (e.g. Ferrier & Pearce 1996). An independent set of validation data for occurrences of *E. globulus* was derived from the atlas of eucalypt distributions where locations were comparable in resolution to the modelling data. Techniques of direct gradient analysis for the presentation of the results of these models as ecological response curves were developed. Maps of the predictions were derived to show not only the mean predicted values, but also the 95% upper and lower confidence intervals, based on the resubstitution analysis of the sample data. These verification analyses provide an indication of prediction error and therefore the potential interpretative applications of the modelled distributions.

The rules of analysis which are essential for robust and precise prediction from the correlation between patterns of distribution and environment are summarised from these applications and discussed in the context of the future potential of ecological response models (Chapter 8). The results provide an indication of the limitations and prospects of correlative models of plant performance. The possibilities for considering other aspects of plant ecology are also discussed.

Box 1.1 A systematic approach to the analysis and interpretation of species' distribution patterns.

1. Compile data that has a minimum set of common attributes for species' response (e.g. presence and absence) and habitat correlates (e.g. location and altitude). [Chapter 2]
2. Assess regional sampling adequacy of compiled data by reference to an ecological classification of landscape attributes, for a scale defined by the study purpose and with levels of sample replication defined by the purpose-set requirements for precision in prediction. [Chapter 2]
3. Define the data subset for the analysis of a particular species' distribution from its potential range or according to the domain in which the sample of presences and absences adequately represent their frequencies in the landscape. [Chapter 3; Williams & Potts 1996]
4. Derive environmental gradients from the main groups of variables — water, temperature, nutrients and light — that are proximal to plant physiological processes. [Chapters 4 & 6].
5. Assess the potential for positive or negative correlations associated with coexisting species that may indicate competitive, neutral or facilitative interactions. [Chapters 5 & 7]
6. Construct multivariate ecological response models from physical attributes of the habitat for potentially competing species. [Chapter 7]
7. Assess the potential significance of biotic attributes with abiotic factors in correlative models of species' occurrence. [Chapter 7]
8. Assess the statistical and ecological validity of the predictive model, and display as a set of geographic distributions and direct gradient responses. [Chapter 7]
9. Assess the potential for interpreting predicted estimates of occurrence from species' distributions in terms of relative performance at a site. [Chapter 7]
10. Evaluate the potential for interpreting species' direct gradient responses in terms of the ecological theory of the continuum concept. [Chapters 7 & 8]

2. Sampling adequacy of compiled ecological data: regional representation of eucalypt forest habitats

2.1 Introduction

This chapter describes the attributes of a set of samples that were compiled for *Eucalyptus* species' occurrences in Tasmania. The objectives for analysis of these data were to predict the distributions of *Eucalyptus* species back onto the landscape using statistical regression procedures and mapping applications. The use of inferential statistics assumes that the samples were selected in a random manner and, as such, can be considered statistically representative of the population or process under study (e.g. McPherson 1990). However, the sample available for studying *Eucalyptus* species' distributions was based on an *ad hoc* collation of data from different sources, for which the sampling methods were not consistently applied, and for which the correspondence between the data and the population of interest cannot be assumed representative. Therefore, this chapter describes the attributes of this sample in the context of an analysis of its representativeness.

The sampling domain of interest to the prediction of *Eucalyptus* species in this thesis is considered as comprising all core and marginal habitats in which eucalypts may grow or persist throughout Tasmania and the Bass Strait Islands. Numerous small and broad scale studies have been undertaken into the distribution and composition of forests in Tasmania. These include broad mapping of vegetation types of which eucalypts are among the most prominent species (Kirkpatrick & Dickinson 1985), and floristic inventories which classify the compositional variability of these forests (e.g. Duncan & Brown 1985; Kirkpatrick *et al.* 1988a). More recently, a finer-scale mapping of all forest communities has been undertaken by generalising existing forest types or direct aerial photograph interpretation (e.g. Hickey *et al.* 1988; Wells 1989; Williams 1989; WGFC 1990). These fine scale mapping studies provide a quantitative overview of the landscape extent of sclerophyll and rainforest vegetation in Tasmania, and may provide a suitable basis for assessing the representation of samples for *Eucalyptus* species' occurrence. However, an assessment of sampling adequacy needs to be firmly established with respect to current theoretical concepts as they relate to the objectives for selecting the sample in the first place.

The simplest approach to assessing the sampling adequacy of compiled ecological data is to undertake a hypothetical vegetation survey within the study area (e.g. Tasmania). Systematic and objective methods for regional vegetation surveys, based on methods of stratified random sampling have been applied in Australia over many decades (e.g. Noy-Meir 1971; Bunce & Shaw 1973; Austin & Basinski 1979). These methods have been extended to take into account the practical difficulties of rugged landscapes, while at the same time directing sampling toward

a representation of plant habitats (e.g. gradient-oriented sampling, or GRADSECTS, Gillison 1984; Gillison & Brewer 1985; Austin & Heyligers 1989, 1991; Margules & Austin 1994). However, no such practical difficulties need be considered for this case of a hypothetical survey design for retrospective assessment of sampling adequacy in compiled ecological data. Nevertheless, the ecological and statistical theory that underpins the GRADSECT design applies to the current analysis.

The ecological theory of field survey design developed for GRADSECT studies is based upon a classification of the landscape that reflects the relationship between plant response and environment. Numerous methods of land classification, also incorporating ecological principles in their derivation, have been developed wherever there is a need for inventory information on which to base decisions of land use (e.g. Novitzki 1995; Bunce *et al.* 1996b; Rowe 1996; Sims *et al.* 1996). For example, land capability surveys aimed at assessing the overall productive potential of land, led to broad classifications of land systems by parent rock type, climate, terrain and vegetation type with an emphasis on soil survey (e.g. Christian 1958; Christian & Stewart 1968; Laut *et al.* 1977; Laut 1981, 1983; Gunn *et al.* 1978; Pemberton 1986; van Overstraeten & Trefois 1993). Other methods of ecological land classification were designed to explicitly examine issues related to conservation and biodiversity (e.g. Thackway & Cresswell 1992, 1995; Belbin 1993; Klijn *et al.* 1995; Smith & Carpenter 1996; Perera *et al.* 1996), or forest and farm management (e.g. Cooper 1995; Bajzak & Roberts 1996; Beauchesne *et al.* 1996; Host *et al.* 1996). In each case, the classification is only as good as its source information, and this is rarely comprehensive for the scale at which predictions, or insightful analyses, are intended (e.g. Sanderson *et al.* 1995; Ryan *et al.* 1995).

For the purpose of predicting species' distribution patterns, using correlative analyses, the ecological classification that is applied to the landscape aims to define a sample which accurately reflects the variability of ecological processes. A representative sample of species' distributions and associated habitats may be viewed as being weighted in proportion to the ecological variability of the landscape, rather than to area. It is the ecological variability that is inherent in the landscape which is the objective of the sampling procedure, rather than the extent of each habitat. Once the variability of habitats are established, for a scale that reflects the purpose of analysis, then replicate samples can be applied within each land class, rather than in proportion of the extent of the land area in each class. Therefore, the initial procedure in the assessment of sampling adequacy is the choice of an appropriate scale for the classification, since this determines the suitability of different information sources with which to address the question of ecological representativeness. This classification scale is defined by the objectives of analysis, since this determines the resolution for which precision in prediction is needed.

A guide to the scale of land units appropriate to different purposes of ecological study and analysis was developed by Klijn & Udo de Haes (1994). Their ecosystems classification nomenclature is reproduced in Table 2.1. Following their nomenclature, a useful scale for

prediction of *Eucalyptus* species' distributions would be equivalent to a map scale of about 1:25 000 to 1:100 000, although for other applications a vegetation mapping scale of about 1:500 000 may also be useful (e.g. Kirkpatrick & Dickinson 1985). These two scales have been termed *ecoseries* and *ecosections*, respectively (Table 2.1). Individual samples for *Eucalyptus* species' occurrences and their habitats are observed in plots of about 0.3 ha, representing an ecosystem classification scale termed *ecotopes* (Klijn & Udo de Haes (994). Therefore, an objective of this analysis of sampling adequacy for eucalypt forest habitats is to define a land classification that represents a scale in the *ecoseries* to *ecosection* range. A representative sample of *Eucalyptus* species' occurrences in *ecotopes* would then be defined in proportion to the ecological variability expressed by the units of that classification, rather than to land area.

Table 2.1 Ecosystems classification scales and nomenclature (after Klijn & Udo de Haes 1994).

Nomenclature	Indicative mapping scale			Basic mapping unit (ha)		
Ecozone	1:	>	50 000 000	>	6 250 000	
Ecoprovince	1: 10 000 000	–	50 000 000	250 000	–	6 250 000
Ecoregion	1: 2 000 000	–	10 000 000	10 000	–	250 000
Ecodistrict	1: 500 000	–	2 000 000	625	–	10 000
Ecosection	1: 100 000	–	500 000	25	–	625
Ecoseries	1: 25 000	–	100 000	1.50	–	25
Ecotope	1: 5 000	–	25 000	0.25	–	1.50
Eco-element	1:	<	5 000		<	0.25

There are practical difficulties to obtaining an appropriate land classification, because of the different purposes involved. Many published classifications are fixed in the definition of class levels and the attributes used in the classification, and may not reflect the attributes or scale required for a particular study (e.g. Mackey *et al.* 1988, 1989). An incomplete or inappropriate classification can lead to spurious patterns (Taylor & Friend 1984). For example, environmental units defined by a broad classification of parent rock type, terrain, and categories of climate would suggest that all subsequent units at this scale were equally variable. But this may not be the case. Other spatio-temporal factors also influence vegetation formation, including disturbance history, local interactions between biota, or patchiness in soil characteristics (e.g. Caswell & Cohen 1991; Williams *et al.* 1994; Houle 1994; Bullock *et al.* 1995; Lathrop *et al.* 1995; Miller *et al.* 1995; Peterson & Squiers 1995; Arnold *et al.* 1996). New flexible approaches to land classification are developing (e.g. Thackway & Cresswell 1992, 1995), but these are still only as good as their source data, and in many regions published classifications still remain the best available source of reference information (e.g. Pemberton 1986). Therefore, variation within and between environmental units needs to be identified wherever existing land classifications inadequately represent the scale or attributes required for predictive modelling (McKenney *et al.* 1996).

Surrogates for ecological variability that account for these differences in diversity or pattern between environmental units can be usefully applied (Faith & Walker 1996). For example, Austin and Heyligers (1989, 1991) and Neldner *et al.* (1995) included geographic extent as an

element in their classifications when devising rules for minimum sampling frequencies. The use of space as a surrogate for ecological variability requires that a model of the relationship between extent and patchiness of units and land area be adopted. By analogy, the species-area curve may be a suitable model for this relationship. Land area is a general reflection of environmental heterogeneity, and the pattern of environmental heterogeneity is a major determinant of the pattern of ecosystem variability (Williamson 1988, 1989). Various mathematical forms for extrapolating the species-area relationship have been examined (e.g. Connor & McCoy 1979; Coleman 1981; Sugihara 1981; McGuiness 1984; Scheuring 1991; Williamson & Lawton 1991; Wissel & Maier 1992; Buys *et al.* 1994). More recently, the extreme-value function (EVF) has been suggested as an appropriate model (Williams 1995, 1996; and see recent application by Mourelle & Ezcurra 1996).

The species-area type relationship may reflect a fundamental property of nature, connecting ecological processes across different scales. For example, the EVF model closely describes the relationship between sampling frequency and land area which was intuitively defined by Neldner *et al.* (1995) to guide the sampling of vegetation units on the Cape York Peninsula (Fig. 2.1). It is also well recognised that different ecosystem processes apply in different regions and at different scales, and this results in different ecological land classifications (e.g. Klijn & Udo de Haes 1994; Sims *et al.* 1996). Therefore, the relationships between environmental heterogeneity and land area, defined by an ecological classification in different regions or at different scales, could be quantified by the parameters of an EVF. These models could then be used to predict ecological variability within and between environmental units of specified size, where an existing land classification incompletely represents the scale or attributes required by the study objectives. In this case, the heterogeneity of environmental units may be used as a surrogate for ecological variability, and applied to the definition of an estimate for minimum sampling requirements within a particular land region.

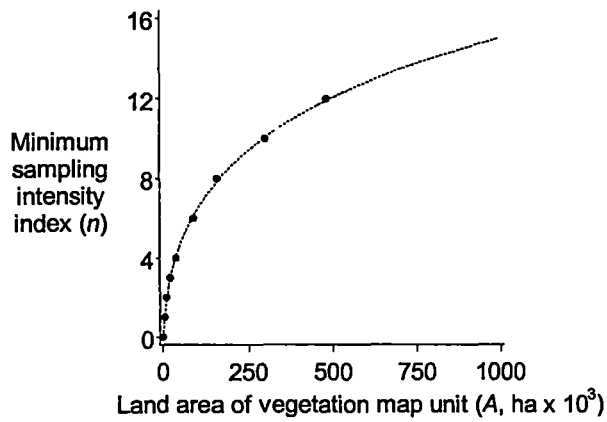


Figure 2.1 Sampling rules devised by Neldner *et al.* (1995) for 1:250 000 scale vegetation survey of Cape York Peninsula. Minimum sampling intensity index (n) by vegetation map unit area (A) is fitted with an extreme value function of the form $n = \alpha\exp(-\exp(\beta-\gamma\log_e A))$, following its application to the species-area relationship by Williams (1995, 1996).

Vegetation patterns can also be used as a surrogate for ecological variability where a land classification incompletely represents the scale or attributes required for representative sampling (e.g. Kirkpatrick & Brown 1994; Carey *et al.* 1995; Bajzak & Roberts 1996; Smalley *et al.*

1996). This is because vegetation properties, such as structure and floristic composition, reflect accumulated variability due to the spatial and temporal history of interactions between species and their environments (e.g. Hobbs & Mooney 1995; Jablonski & Sepkoski 1996; Moloney & Levin 1996; Wilson *et al.* 1996). These effects accumulate at broader scales and emphasise differences in ecological variability observed between geographic regions (e.g. Pacala & Deutschman 1995; Ferrier & Watson 1996; He *et al.* 1997). However, sampling by vegetation map units only is unlikely to ensure a representative sample of ecological variability (e.g. Neldner *et al.* 1995), because contrasting environmental factors may result in similar vegetation patterns. The resulting sample will be limited to a subset of possible interactions that may be location-specific and less robust in predictive models. Therefore, vegetation features are a useful surrogate for unknown sources of ecological variability, only when defined independently of other landscape features of the physical environment that are already included in the classification.

In seeking to address the sampling adequacy of compiled data for Tasmania, I found that existing published land (e.g. Pemberton 1986) and vegetation classifications (Wells 1989; Williams 1989) for eucalypt forest habitat were defined for relatively broad scales or did not adequately reflect the key attributes needed to distinguish regional ecological processes. For example, the land systems classification for Tasmania (e.g. Pemberton 1986), represents land units at the *ecosection* scale to *ecodistrict* scale (about 25 ha to 10 000 ha), but there is no distinction of vegetation type and the published reports that list these areas represent even broader scales (between *ecosection* and *ecoregion*) than the original 1:100 000 scale survey maps. A second landscape classification was specifically directed toward summarising eucalypt forest habitat, but only considered a broadly defined classification based on the generalisation of geology and altitude classes across the land systems, within biogeographic regions (e.g. Wells 1989; Williams 1989). The combination of these two information sources would represent a reasonable classification scale (in the *ecosection* range) for assessing the sampling adequacy of eucalypt forest habitat in the compiled data. However, the two land classifications can only be approximately matched across biogeographic regions, rather than between classification units within regions.

This chapter therefore develops a statistical approach to combining the published classification information in the land systems and the eucalypt forest reports. This approach is soundly based in species-area theory and associated statistical matching between land area and environmental heterogeneity (e.g. Fig. 2.1). The environmental classification of the land systems was used to define a relationship between land area and environmental heterogeneity in each of the seven reported biogeographic regions for Tasmania (e.g. Pemberton 1986). These regional relationships were then used to estimate expected levels of environmental heterogeneity within each of the classification units for eucalypt forest habitat that could be generalised between the two reports for sclerophyll forest occurrence (Wells 1989; Williams 1989). These levels of environmental heterogeneity were then subjectively weighted (e.g. Fig. 2.1) by a relationship based on land area

to account for surveyor-determined differences in mapping intensity reported for the land systems in different biogeographic regions.

A direct relationship between the expected levels of environmental heterogeneity and expected levels of sampling for representing anticipated levels of ecological variability within and among eucalypt forest habitat units was then defined. Replicate samples could then be determined in accordance with the objectives for predictive precision. A simple chi-squared comparison could then be used to test the significance between observed and expected levels of sampling, providing a statement of sampling adequacy, sampling bias, and ultimately suggesting options for reviewing either the objectives of analysis or the requirements for supplementary sampling.

An additional analysis was undertaken to explore the level of ecological heterogeneity that was inherent to the ecotope sampling units. Rarefaction methods (e.g. Heck *et al.* 1975; Colwell & Coddington 1994) were also used to derive a species-area type relationship, but this time between sampling intensity and ecological variability at the *ecoseries* scale, defined by a classification of the biotic and abiotic attributes of each ecotope (i.e. the plot samples of about 0.3 ha each). Therefore, the classifications for either biotic or abiotic heterogeneity were individually assessed and extrapolated to an asymptote, providing both an estimate of potential heterogeneity and the sampling requirements needed to achieve an *ecoseries* scale resolution in subsequent predictive analyses. The separate classifications of biotic and abiotic heterogeneity also enabled local and regional differences in ecological variability to be compared.

The results of these analyses provide an overview of the ecological variability of eucalypt forest habitats in Tasmania, and an assessment of the regional and local sampling adequacy of a compiled set of ecological data for predicting *Eucalyptus* species' occurrence in Tasmania.

2.2 Methods

2.2.1 Description of study area

The study area is the Tasmanian archipelago (*c.* 6 850 000 ha), off the coast of South-eastern Australia (*c.* latitudes 40 to 43.5 °S, longitudes 144 to 148 °E). It comprises a large continental island, and smaller coastal and Bass Strait islands. The southerly location accentuates seasonal variability in day-length and climate, and the interaction with topography creates a large diversity of environments with locally steep climatic gradients. The climate is temperate–maritime with a slight continental influence. Coastal dunes, hills and plains, give way to broad glacial and inland river valleys, mountain peaks and plateaux. Coastal daily temperatures range by about 7°C and are almost double this inland (see climate statistics in Bureau of Meteorology 1993). Mean annual rainfall varies from about 500 mm in the rain-shadowed inland areas of the Midlands and coastal areas of the east and south-east, but may exceed 3200 mm in western mountain regions. Mean temperatures range from a minimum of 2 to 6°C in the winter months, and a maximum of 18 to 23°C in the summer months; with lower averages at higher elevations. Inland frosts occur

all year round, being more frequent in the winter months and in situations of cold air drainage. Snow falls on the higher peaks in the winter months, localised droughts and floods are not uncommon, and westerly winds are particularly persistent during late winter to early spring.

The geology further contributes to the diversity of environments; and sharp boundaries lead to the occasional juxtaposition of distinct habitats. For example, the extremes in nutrient levels of soils derived from serpentinite rocks may lead to the local development of open sclerophyllous scrub next to tall closed rainforest (e.g. Gibson *et al.* 1992). In general, landscapes in the east are dominated by Jurassic dolerite, Permian, and Triassic (Permo-Triassic) sediments, and in the West, older Precambrian and Cambrian sediments and metamorphics prevail (Department of Mines 1976). The general coincidence between climate, land-forms and substrate emphasises broad biogeographic differences between western and eastern environments in Tasmania (e.g. Lewis *et al.* 1991; Kirkpatrick & Brown 1991), and have contributed to the delineation of biogeographic regions (e.g. Orchard 1988; Thackway & Cresswell 1995).

The vegetation of Tasmania is structurally and floristically diverse, ranging from closed forest, woodland and scrub to moorland, heath and grassland, and from lowland to montane habitats (Kirkpatrick & Dickinson 1984). The floristic composition of broad groupings of forest type has been classified for rainforest (Jarman *et al.* 1984; 1991), swamp forest (Pannell 1992), wet eucalypt forest (Kirkpatrick *et al.* 1988a), and dry sclerophyll forest (Duncan & Brown 1985). The conservation status of these forest communities has been summarised with other classes of vegetation by Kirkpatrick *et al.* (1995). The extent of these broad forest types was mapped and environmentally classified by Hickey *et al.* (1988), Wells (1989) and Williams (1989) as part of a quantitative assessment of their conservation values (Hickey & Brown 1989; WGFC 1990).

2.2.2 Compiled ecological data — a biophysical sample of eucalypt occurrence

An ecological dataset comprising 15 611 geo-referenced samples (0.1 to 0.3 ha plots) of predominantly eucalypt-dominated forests (14 980 observations), describing their physical and biological environments, was made available for predictive analyses by Forestry Tasmania. These data consist predominantly of the continuous forest inventory (CFI) records for forests of production potential (Lawrence 1978), comprising a serious of temporary and permanent plots, for which the latter were collated for the most recent monitoring period up to November 1994. These CFI data were supplemented by dry sclerophyll forest plots (Duncan & Brown 1985), assessment surveys of recommended areas for protection (Williams 1989), selected pre-logging surveys of National Estate listed State forest (Forestry Commission Tasmania unpublished data, 1987 to 1989), and other survey data from particular ecological or genetic studies (Table 2.2).

Table 2.2 Sources of collated information, the number of observations in each case (*n*), and some associated publications.

n	Source	Institution	Reference
Continuous Forest Inventory.			
2668	Forestry staff	Forestry Tasmania (permanent plots)	Lawrence 1978
8861	Forestry staff	Forestry Tasmania (temporary plots)	Lawrence 1978
Dry Sclerophyll forest survey·			
2564	F. Duncan	National Parks and Wildlife Service / Tasmanian Forestry Commission	Duncan 1981, 1983, 1986 Duncan 1988, 1989a, b Duncan *et al.* 1981 Duncan & Harris 1983 Duncan & Duncan 1984 Duncan & Brown 1985 Duncan & Williams 1988 Kirkpatrick & Duncan 1987 Brown & Duncan 1989
197	M. J. Brown	Tasmanian Forestry Commission (National Parks and Wildlife Service)	Brown & Bayly-Stark 1979a, b Brown & Buckney 1983
569	F. Duncan (various rare species records collated by)		Duncan 1988 Kirkpatrick 1983 Kirkpatrick *et al.* 1980 Kirkpatrick & Brown 1984a, b
National Estate and Recommended Area for Protection surveys·			
544	K. J. Williams	Tasmanian Forestry Commission	Unpublished data 1989 Williams 1989 Williams & Duncan 1991
Other Survey data:			
121	S. Harris	Dept. Parks, Wildlife and Heritage	Harris 1987, 1989 Harris & Brown 1980 Harris & Kirkpatrick 1991
40	R. Wiltshire	Dept. Plant Science (Uni Tasmania)	Wiltshire *et al.* 1991a, b, 1992
38	B Potts	Dept. Plant Science (Uni. Tasmania)	Potts 1988, 1989 Potts & Reid 1985a, b
29	J. B. Davies	Dept. Agriculture	Davies 1988a, b
18	M. Savva	Dept. Plant Science (Uni. Tasmania)	Savva *et al* 1988
11	M. Neyland	Dept. Parks, Wildlife and Heritage	Neyland & Duncan 1988
4	N. Gibson	Tasmanian Forestry Commission	Gibson 1987
4	J. C. Grant	Tasmanian Forestry Commission	Unpublished data 1988

The floristic data in common between these sources consisted of the presence or absence and rank order dominance of *Eucalyptus* species and other tree species that distinguished sclerophyll from rainforest vegetation types. Where field observations were not available for forest structure and cover classes, these were derived from 1:25 000 scale, photo-interpreted forest type maps (for interpretation methods see Sulikowski 1995).

The environmental data collated with these samples included geology, altitude, aspect and slope. Missing values for geology were derived from the 1:500 000 Geological Map of Tasmania (Department of Mines 1976). Missing values for altitude, slope and aspect were estimated from 1:100 000 TASMAP series (less than 3% of the data). Additional environmental information that would be useful for predictive modelling (e.g. drainage, topographic position, disturbance and fire frequency) was not consistently recorded between sources (Williams 1990). An index of potential nutrient supply from Tasmanian parent rock types, on a 10 point scale, was derived from the table given in Nix *et al.* (1992). This index defines the mean opinion of Tasmanian forest soil scientists (B. Nielson, J. Honeysett, M. Pemberton, J. Grant) based on the 36 lithology classes defined and delineated by the 1:500 000 geological map (Department of Mines 1976). The lowest values of 2.0 for Tasmania represent the shallow to deep highly organic peats

developing on Cambrian and Precambrian orthoquartzite or metamorphic rocks of predominantly metaquartzite (Grant *et al.* 1995). The highest values were 8.9, representing the shallow and stony clay loams or deep, rich, red-brown gradational soils on Tertiary basalt (Grant *et al.* 1995).

Monthly estimates of climate (maximum and minimum temperature, precipitation, rain-days, potential evapotranspiration, flat-surface and terrain-modified solar radiation) were estimated from the topographic (aspect, slope) and location (latitude, longitude, altitude) data using the process model ESOCLIM (Version 3.6, May 1995, developed by H. A. Nix, J. R. Busby, M. F. Hutchinson & J. McMahon; see User's Guide, McMahon *et al.* 1996). The spatial variation of each climatic parameter is represented by a mathematical interpolated surface. Values are derived by resolving the surface coefficients at a specified location. Although the climatic surfaces are continuous, the coefficients themselves can be resolved on a regular grid of elevations to produce gridded estimates of the climatic parameters. For Tasmania, the minimum geographical area that these surfaces are discriminating reflect a grid resolution of 0.01 minutes. This approximates an area resolution of 200 by 200 m or 4 ha.

2.1.1 Information sources for ecological land classification

2.1.1.1 Land systems

The land systems of Tasmania are a fixed-scale classification of key landscape attributes that were subjectively mapped following the review of existing information and extensive field surveys in seven geographic regions: (i) King Island, Richley (1984); (ii) Flinders Island, Pinkard & Richley 1984; (iii) North West, Richley (1978); (iv) North East, Pinkard (1980); (v) Central Plateau, Pemberton (1986); (vi) South, East & Midlands, Davies (1988a); (vii) South West, Pemberton (1989). The 462 land systems are the basic spatial units of this classification. These were defined from regional combinations of nine classes of rainfall, 45 geological types, 300-m altitude zones, six topographic types, and with additional variation in soils or vegetation to further distinguish similar land units. The land systems were mapped at 1:100 000 scale, but the published reports summarise the extent and major features of each land system, effectively amalgamating the spatial heterogeneity of constituent patches of the same type within a region. The mapped representations of the land systems reflect ecosystem processes at the *ecosection* scale, but the reported extents summarise this inherent variability at a broader classification scale: *ecosection* to *ecodistrict*.

2.1.1.2 Eucalypt forest habitat areas

The extent and distribution of sclerophyll forest vegetation (eucalypt dominated) were also mapped at 1:100 000 scale (WGFC 1990). Intersection of these areas with an amalgamated land systems classification of geology and altitude classes provided a basis for distinguishing differences in vegetation-by-environment combination in Tasmania. The extent and tenure of these 'habitat units' were used to assess the conservation status and reservation requirements of the respective forest types (Wells 1989, Williams 1989). These published reports summarise the spatial heterogeneity of these

habitat units, and indicate the extent of each within classes of geology and altitude within biogeographic region. Similar to the land systems, the mapped representations of these 'eucalypt forest habitat units' reflect ecosystem processes at the *ecosection* scale, but the reported extents summarise this inherent variability at the broader classification scale of *ecosection* to *ecodistrict*.

The biogeographic regionalisation was based upon the 'Forestry classification' reported by Orchard (1988). This classification delineates 12 different regions, whereas the land systems regionalisation distinguishes seven regions. Some of these regions share boundaries (e.g. King and Flinders Islands), others share similar environments and approximate boundaries. For example, the 'North West' region defined for the land systems (Richley 1978) largely includes the 'Northwest' and 'North & Northern Midlands' biogeographic regions (Orchard 1988), and the 'South, East & Midlands' land system region (Davies 1988a) generally corresponds to the 'East & Southern Midlands' and 'South Dolerite' biogeographic regions (Orchard 1988).

Since the level of environmental classification of these forest assessments differed, the broader classification of Wells (1989) was used to derive a combined dataset of eucalypt forest habitat units.

The resulting classification of eucalypt forest habitat units is distinguished by eight biogeographic regions:

King Island, Furneaux Islands, Northwest, North & Northern Midlands, Northeast Lowlands & Highlands, East & Southern Midlands, Centre, South Quartzite & West, South Dolerite;

fourteen geology groups:

Precambrian metamorphics (PM), Precambrian sediments (PS), Cambrian volcanics (CV), Cambrian sediments (CS), Ordovician quartzites (OQ), Ordovician limestone (OL), Devonian granite (DG), Mathinna beds (MT), Permo-Triassic sediments (PT), Jurassic dolerite (JD), Tertiary basalt (TB), Tertiary sediments (TS), Quaternary till & talus (QT), Quaternary deposits (QD);

and three altitude classes:

Sea level to 600 m, 600 m to 900 m, 900 m to the tree-line (about 1200 m).

2.1.1.3. Local-scale variation in eucalypt habitats

The ecological variability inherent to the local plot samples of eucalypt forest (c. 0.1 to 0.7 ha, median area ~ 0.28 ha). Since each sample reflects a unique record of vegetation by environment combinations, a classification was derived to represent an estimate of ecological variability approximated to an ecosystem classification of *ecoseries* (i.e. up to c. 25 ha of amalgamated patches, irrespective of inherent spatial heterogeneity). This was achieved by considering the important components of ecosystem variability related to indices for water, temperature, light, nutrients, and forest structure and composition, attributes for which were available within the ecological dataset. To retain a separate consideration of physical and biological variability at the *ecoseries* scale of classification, the two sets of attributes were independently classified (Table 2.3).

Table 2.3 Classification of *ecoseries* from the biotic and abiotic attributes recorded in plot samples of eucalypt forest occurrence. The number of samples (*n*) in each category of are shown with the class limits, class mid-points and class-mean values from the ecological dataset (*n* = 15611).

Water is the difference between annual precipitation and annual potential evapotranspiration (mm/year); Temperature = mean annual minimum temperature (°C); Light = mean annual terrain-adjusted solar radiation (MJ m^{-2}); Inclination = terrain slope (°).
Six classes of understorey type: (i) *Casuarina* or *Callitris* (open) forest or non-forest: heathland, scrub or moor, HS; (ii) Dry sclerophyll eucalypt (open) forest: understorey of sclerophyllous shrubs, heaths, or grasses, DF; (iii) Wet sclerophyll eucalypt forest: understorey of broad-leaved shrubs or wet ferns, WF; (iv) Eucalypt-rainforest (closed forest): understorey of rainforest tree-species, MF; (v) Wattle forest (closed forest): *Acacia melanoxylon* or *A. dealbata* canopy, BF; (vi) Rainforest (closed forest): eucalypts absent, RF.
From the published accounts, 24 *Eucalyptus* species may occur as forest dominants (Duncan and Brown 1985; Kirkpatrick *et al.* 1988a).

Abiotic attributes						Biotic Environment					
ID	> min	≤ max	mid	mean	*n*	ID	> min	≤ max	mid	mean	*n*
Water:						*Eucalypt presence or absence:*					
1		-250	-450	-358	1488	1	Present				14952
2	-250	-50	-150	-144	2402	2	Absent				659
3	-50	150	50	48	2822						
4	150	350	250	248	2992	*Eucalypt dominance:* 24 classes					14952
5	350	550	450	448	2716						
6	550	750	650	641	2048						
7	750		1300	931	1144	*Eucalypt species richness (count)·*					
						1	1				6695
Temperature:						2	2				5423
1		3.0	1.75	2 4	638	3	3				2253
2	3.0	4.0	3.50	3.6	1211	4	4				507
3	4.0	5.0	4.50	4.6	2109	5	5				61
4	5.0	6.0	5.50	5.6	4141	6	6				9
5	6.0	7.0	6.50	6.5	3695	7	7				4
6	7.0	8.0	7.50	7.4	2338						
7	8.0	9.5	8.75	8.5	820	*Average forest stand height (m).*					
						1		5	2.5	4.2	482
Nutrient Index:						2	5	15	10.0	11.7	338
1		3.50	3.000	3.33	2571	3	15	25	20 0	20.4	3123
2	3.50	4.00	3.750	3.66	1331	4	25	35	30.0	33.4	4292
3	4.00	4.50	4.250	4.47	2297	5	35	50	42.5	46.3	5515
4	4.50	4.75	4.625	4.69	434	6	50	65	57.5	65.0	1202
5	4.75	5.00	4.875	4.81	2759						
6	5.00	7.00	6.000	5.69	545	*Average forest stand cover (%):*					
7	7.00	9.00	8.000	7.75	5015	1		5	2.5	5.0	857
						2	5	15	10.0	12.3	897
Light:						3	15	25	20.0	20.0	794
1		4100	3500	3829	756	4	25	50	37.5	33.2	5019
2	4100	4500	4300	4342	1725	5	50	75	62.5	57.3	3475
3	4500	4700	4600	4610	2031	6	75	100	87.5	80.5	3910
4	4700	4900	4800	4798	2538						
5	4900	5100	5000	5007	2978	*Understorey type:*					
6	5100	5500	5300	5276	4938	1	HS				215
7	5500	6100	5800	5581	646	2	DF				6449
						3	WF				6793
Inclination:						4	MF				1710
1	0	10	5	4	8692	5	BF				315
2	10	20	15	14	4591	6	RF				129
3	20		30	27	2328						

Abiotic heterogeneity of *ecoseries* was derived from a classification of five indices: (i) climatic water balance (difference between total annual precipitation and evaporation), (ii) average temperature limits (mean annual minimum temperature), (iii) the realised light environment (cloudiness and terrain-modified mean annual solar radiation), and (iv) a topographic position index based on three classes of terrain slope: < 10°, 10-20°, > 20°, representing generally flat topographies, gently rolling hills, and steep mountain slopes, hillsides or cliffs.

Biotic heterogeneity of *ecoseries* was derived from a classification of six forest structure and composition classes: (i) presence or absence of eucalypts, (ii) the dominant *Eucalyptus* species of a stand, (iii) *Eucalyptus* species' richness, (iv) maximum estimate of forest stand height (e.g. old growth class, Sulikowski 1995), (v) maximum estimate of forest stand cover, (vi) indicative forest tree and understorey life forms that suggest vegetation type in six classes: rainforest, blackwood forest, mixed forest (sclerophyll-rainforest), wet sclerophyll forest, dry sclerophyll forest and woodland, or non-forest vegetation classes including moorland, heathland and grassland from lowland to alpine environments.

2.2.4 Quantifying ecological variability

2.2.4.1 Regional to local scale environmental heterogeneity

This assessment of sampling adequacy requires a classification of ecological variability for eucalypt forest habitat units at the *ecoseries* to *ecodistrict* scale. However, the existing classification for eucalypt forest is not adequate for representing key processes at this scale. A suitable classification needs to include key ecological attributes for the physical (water, temperature, nutrients, and light) and biological (forest type, structure and composition) environments. The land system classification includes more information about the physical environment, but does not distinguish forest type. In combination, the attributes for the two classifications provide surrogates for the key attributes believed to be important to the variability of these eucalypt forests. For example, within the regional distributional extent of eucalypt forest, the water regime could be inferred from region and altitude interactions, temperature regimes from altitude and region interactions, nutrients status from geological type and region interactions, and light regimes from region and topography interactions. A means of combining the two sets of information from the land systems and the eucalypt forest habitat units was therefore needed.

An analysis of land system heterogeneity within each land system region was considered a suitable basis for approximating the environmental heterogeneity within comparable land areas for comparable regions of the existing classification for eucalypt forest habitat units. The regional relationships between land system heterogeneity and land area were based on an extension of species-area theory to ecological variability, since environmental heterogeneity is believed to be a major driving force behind the landscape diversity of species, communities and ecosystems (e.g. Williamson 1988, 1989). The derivation of a relationship between land system diversity and

land area in each region was therefore considered as a reasonable estimate of the *environmental heterogeneity* in the respective region.

Models of environmental heterogeneity were developed from the cumulative diversity of land systems with area in each region. The land systems in each region were randomly selected (without replacement) in units of 110 ha, reflecting the size of the smallest land system, to construct each environmental accumulation curve in an unbiased manner. To overcome the large number of 110 ha patches in some regions, points on the curves were defined from summaries of mean land system frequency for data grouped by the power series of 2^i ($i = 2, 4, 8, ..., n$), and the total number of land-area samples in each case (n). Since the order of sample selection was also important, 100 randomised resamplings were undertaken (e.g. Fig. 2.2A).

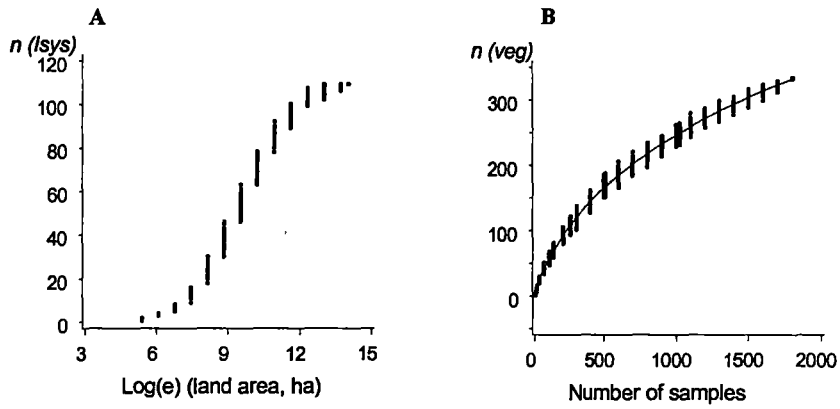


Figure 2.2 Example of observed data spread for 100 randomised resamplings of regional land system heterogeneity in the North-East (A) and local scale biotic heterogeneity on Jurassic dolerite in the South region (B). For the analysis of land system heterogeneity in the North-East region [transcribed from Pinkard (1980)], data were grouped by the power series of 2^i ($i = 2, 4, 8, ..., 8192$), and the total sample size ($n = 11\,301$ units of 110 ha). For the analysis of scale biotic heterogeneity on Jurassic dolerite in the South region, data were grouped by the power series of 2^i ($i = 2, 4, 8, ..., 1024$), by units of 100 (i = 100, 200, 300, ..., 1800) and the total sample size ($n = 1808$ observations). The predicted heterogeneity from the fit of an extreme value function is also shown for the local scale analysis (B).

Resampling adequacy was assessed by examining the scree curve for the reduction of the standard error to mean ratio (after Dale *et al.* 1991). The standard error to mean ratio declined to negligible levels (< 10%) within 50 to 100 resamplings for all data groups. The smaller data groups (i.e. less than about 128 land area samples of 110 ha patches) required the larger frequency of resamplings and therefore set the statistical minimum level at 100 randomised resamplings. Following an extension of the logic for the use of the extreme value function (EVF) in cases of the species-area relation (Williams 1995, 1996), a curve of the Gompertz form (Ratkowsky 1990) was fitted to the resampled sets of randomly accumulated environmental heterogeneity (Box 2.1). The maximum number of land systems in each region defined the asymptote of these relationships. Other parameters which characterise the average form of each environmental accumulation curve, and its upper and lower confidence intervals, were estimated using the NLIN procedure in SAS (SAS Institute Inc. 1990c).

Box 2.1 Derivation of the extreme value function (EVF) model for the relationship between multi-species and area (after Williams 1995) and its application to the relationship between environmental heterogeneity and land area.

A candidate derivation for an approximation to the multiple species-area relation by Williams (1995) is the extreme value function model:

$S = P\{1\text{-}\exp(\text{-}\exp(y\log A + \log d))\}$

where:

S is the expected number of species within a source region and is the sum of the incident functions of each species with an approximate binomial distribution.
P is the participating number of species in the source region (the 'participating' size of the source pool being those species that occur in the observed data; being less than the 'potential' pool-size which is composed of all those species occurring within the source region).
A is the area of the source region.

Generalisation of the model is provided by the parameters where:
$\log d$ allows the intercept to vary; and y allows the slope to vary.

Similarly, the relationship between multiple environments and sample size or area may be defined as:

$E = P\{1\text{-}\exp(\text{-}\exp(y\log A + \log d))\}$

where E is the cumulative number of different environments, species, or spatial units of interest,
A is the cumulative sample size or area, and
P is the participating number of environments known to exist;
all within the source region.
In this case, the participating number of environments is taken as the actual number, which depends upon the scale and attributes of the ecological land classification used for the source region.

In the Gompertz form (after Ratkowsky 1990) the EVF model defines the relationship between multiple environments (E) and sample size or area (A) as:

$E = P\{1\text{-}\exp(\text{-}\exp(\beta - \gamma \log_e A))\}$

where the parameters β and γ are related to the EVF model of Williams (1995) as:
$\log d = \beta$ and
$y = -\gamma$.

Since the classification of land systems by different surveyors resulted in some extensive areas (e.g. Davies 1988a) that were broader than the general *ecosection* to *ecodistrict* scale. A set of rules, similar to those devised by Neldner *et al.* (1995) and Austin & Heyligers (1989), were conservatively used to adjust estimates of environmental heterogeneity by retrospectively incorporating land area in the stratification as a surrogate for undefined factors of ecological variability (Table 2.4).

Table 2.4 Spatial stratification rules for undefined levels of spatio-temporal heterogeneity within eucalypt-forest habitat units.

Area of habitat unit (ha)			Weighting of heterogeneity
	<	25 000	1
25 001	–	50 000	2
50 001	–	100 000	3
100 001	–	200 000	4
200 001	–	400 000	5

While the EVF models were a close fit to the observed trend in environmental heterogeneity, there was a consistent pattern in the residuals. The models slightly over-predicted the number of land systems for small areas relative to the total area of the region, particularly below about 500 ha to 1800 ha, depending upon the region. This over-prediction of average cumulative land system frequency, although small in comparison to the overall fit of the curve, may influence the

accuracy of estimating sampling requirements in habitat units of comparably minor extent. For example, a rare habitat type for eucalypt forest on Tertiary basalt in the 600 to 900 m altitude range from the North-East region has an area of 180 ha for which up to four land systems were predicted, but an estimate of two is closer to reality. Without finer scale information to clarify these estimates, and for consistency with later predictions, scaled adjustments to these few, small areas were made according to the differences between observed and predicted number of land systems.

2.2.4.2 Local scale biotic and abiotic heterogeneity

The ecological variability inherent to the existing sample of eucalypt forest for a classification scale of *ecoseries* within specified habitat units was examined using a similar approach based on species-area theory (e.g. Williamson 1988, 1989). In this case, actual levels of ecological variability were unknown and could only be inferred by extrapolating from the respective classification derived from each set of samples. This approach taken therefore represents an application of rarefaction (e.g. Heck *et al.* 1975; Birks & Line 1992; Colwell & Coddington 1994). The respective classifications of biotic and abiotic heterogeneity were considered separately. The extrapolated asymptotes of the respective rarefaction curves were taken as an estimate of *ecoseries* scale biotic and abiotic (environmental) heterogeneity within the eucalypt-forest habitat units. Sampling adequacy was determined relative to the asymptote for each rarefaction analysis of biotic and abiotic heterogeneity. The different relationships between biotic and abiotic heterogeneity were also used to compare average trends in ecological processes between habitat units.

Four well-sampled eucalypt forest habitat units ($n > 900$), grouped across their constituent altitude classes, were selected for demonstrating the rarefaction analysis of biotic and abiotic heterogeneity (Table 2.5). This selection of habitat units was designed to maximise the proportion of *ecoseries*-scale environmental heterogeneity likely to be represented by each set of observations. For the purpose of characterising the environmental accumulation curves, the data were grouped by the power series of 2^i ($i = 2, 4, 8, ..., n$), and also by units of 100 (i.e. 100, 200, 300, ..., n) up to the total number of observations in each habitat unit (n). No area-weighting need be applied since each sample was approximately equivalent in size (i.e. 0.1 to 0.3 ha plots) and was assumed to represent the same basic ecological unit. The standard error to mean ratio declined to negligible levels (< 10%) within 100 resamplings for all data groups in each case, except for the accumulation of abiotic heterogeneity of Permo-Triassic sediments in the 'South' region (standard error to mean ratio declined to about 17.5% following 100 resamplings). The 10% rule for reduction in the standard error to mean ratio for assessing statistical adequacy of a sampling regime (after Dale *et al.* 1991) indicated that more than 100 resamplings of a randomised order of accumulation may be required to accurately characterise some curves when extrapolating local scale environmental heterogeneity from an existing set of biophysical observations.

Table 2.5 The number of different environments resulting from a local scale classification of abiotic [*n*(Abiotic)] and biotic [*n*(Biotic)] attributes for a sample of ecological variability in four habitat units. The area and number of samples [*n*(sample)] in each habitat unit are also shown.

Habitat unit which defines the spatial domain				Classification	
Region Geology	Altitude	Area (ha)	*n*(sample)	*n*(Abiotic)	*n*(Biotic)
South					
Permo-Triassic sediments	Sea-level to 1200 m	75 685	985	310	172
Jurassic dolerite	Sea-level to 1200 m	155 230	1808	438	334
North-East					
Mathinna beds	Sea-level to 900 m	134 105	1505	376	330
Devonian granite	Sea-level to 1200 m	165 020	1635	469	318

As in the regional case, the parameters and associated confidence intervals characterising the local scale environmental (biotic or abiotic) heterogeneity, were estimated from the average fit of the Gompertz model (Box 2.1) to the 100 randomised resamplings (e.g. Fig. 2.2b). To estimate biophysical sampling adequacy of ecological variability, the asymptote for the EVF model was initially defined from the observed maximum number of environments for each biotic or abiotic classification, and iteratively increased until the residual mean square and pattern associated with the residuals was minimised.

Trends in ecological processes within each habitat unit were estimated from the parameters of a linear regression in which biotic heterogeneity is the response and abiotic heterogeneity is the explanatory variable. These regression equations were compared among the four habitat units.

2.2.5 Assessing sampling adequacy in eucalypt-forest habitat units

The regional relationships between land system diversity and land area defined average levels of environmental heterogeneity with respect to the specified land extents of each eucalypt forest habitat unit, within the comparable biogeographic region. These estimates were adjusted by the spatial stratification rules (Table 2.4) to account for inconsistencies between the land system classifications. The subsequent estimates of environmental heterogeneity for each eucalypt forest habitat unit were then directly related to expected requirement for minimum representative sampling (without replication), assuming spatially-random allocation of samples within each habitat unit. However, replicate samples are also a statistical requirement for analysis. Therefore, a minimum of two to three samples for each predicted level of heterogeneity might be needed. The number of replicate samples depends upon the precision required in prediction of ecological processes within the selected classification scale (e.g. *ecoseries* to *ecodistrict*). Ultimately, however, this depends upon the original purpose and resources available to the field survey, or subsequent sample compilation. Conversely, the classification scale at which minimum replication is currently achieved could be used to guide an assessment of the scale of ecological processes at which reasonable precision in prediction might be expected.

In the current study, the ecological dataset comprises about 15 000 biophysical observations. This sets the target for determining replication levels when deriving expected sampling

frequencies. The observed samples were grouped according to the same environmental land classification used to define the eucalypt forest habitat units (defined in Section 2.2.3.2). Their sampling adequacy was subsequently assessed by a Chi-squared test of differences between observed and expected sampling frequencies in each eucalypt forest habitat unit, following the log-likelihood method of McPherson (1990). Where there was no significant difference ($p > 0.05$), the observed sample was considered to adequately represent ecological variability at the *ecosection* to *ecodistrict* level of classification. This is the limiting scale set by the resolution of land information available from the published accounts of the land systems (e.g. Pemberton 1986) and the amalgamation of habitat information between the sclerophyll forest reports (Wells 1989; Williams 1989).

At the *ecoseries* scale, expected sampling frequencies were inferred from the extrapolation of the environmental accumulation curve to an asymptote (defined by the fit of the extreme value function model to the sample data) for each of four habitat units (indicated in Table 2.5). To assess relative sampling adequacy, the expected number of randomly located samples required to obtain close to the maximum number of environments (predicted by the asymptote of the EVF model) was compared with the observed sample size.

2.3 Results

2.3.1 Land systems and eucalypt forest habitat units

The average size of land systems in each region differs (Table 2.6); reflecting real differences in ecosystem diversity and extent between regions as well as variations in the accuracy of source data and in each surveyor's weighting of source data. Nevertheless, the location-accuracy of the geology and altitude attributes were emphasised when drawing the map boundaries (Pemberton 1986), supporting their use as the environmental basis for defining eucalypt forest habitat diversity (Hickey & Brown 1989; WGFC 1990).

Table 2.6 Summary of land systems by regions in Tasmania. Total number of land system, average regional area of land systems, and proportion of land systems represented in each region (excluding lakes).

Land System Region	Land area (ha)	land system (frequency)	Mean area (ha)	land system (%area)
King Island (Richley 1984)	135 349	16	8495	3.46
Flinders Island (Pinkard & Richley 1982)	173 635	18	9646	3.90
North West (Richley 1978)	1 594 461	93	17145	20.13
North East (Pinkard 1980)	1 248 570	110	11351	23.81
Central Plateau (Pemberton 1986)	677 710	55	12322	11.90
South, East & Midlands (Davies 1988a)	1 752 069	130	13477	28.14
South West (Pemberton 1989)	1 101 122	40	27528	8.66
Totals	*6 682 916*	*462*		

Eucalypt forests are extant (*c.* 1985) across 37% of the land area in Tasmania. Of this area, eucalypts occur more in low forests (59%, predominantly 'dry sclerophyll', Williams 1989) than in tall forests (41%, predominantly 'wet' types, Wells 1989). Low forests and open woodlands predominate in the dry lowlands and the subalpine habitats of northern, north-eastern, eastern and midland regions (Duncan & Brown 1985). Tall forests predominate in the cool to warm, moist habitats of western, north-western, and southern regions (Kirkpatrick *et al.* 1988a). These climatic associations correspond to a trend in substrate differences in which tall forest types predominate on the more mesic habitats associated with older Precambrian, Cambrian, Ordovician and Tertiary type substrates in western-regions (Wells 1989), and low forests predominate on the younger Quaternary deposits, Permo-Triassic sediments, Mathinna beds, Devonian granite and Jurassic dolerite of the upland or drier, eastern regions (Williams 1989).

These eucalypt forest vegetation types comprise 110 habitat units when defined by combinations of eight biogeographic regions, 14 geological groups and three altitude classes. These habitat units vary in size from 180 ha (highland Tertiary basalt), in the Northeast region, to over 382 000 ha (lowland Jurassic dolerite), in the East & Southern Midlands region. Most habitat-units, however, exist across a land area ranging from about 1000 ha to 10 000 ha (Fig. 2.3), representing a classification of ecosystems between *ecosection* and *ecoregion*.

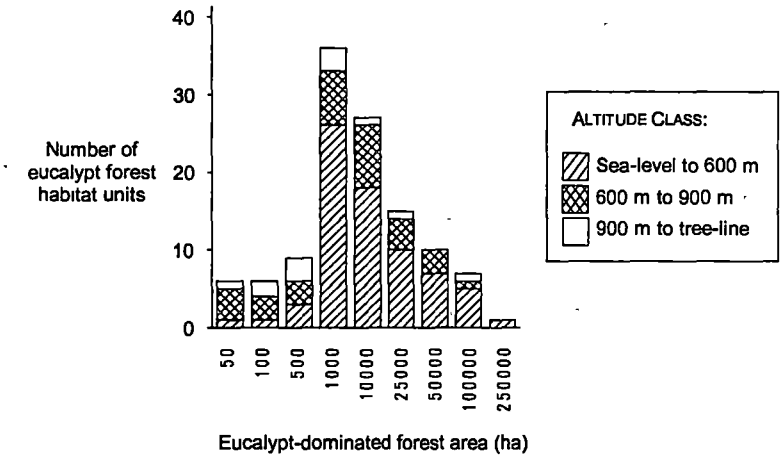


Figure 2.3 Size range of eucalypt forest habitat units in Tasmania (after Williams 1989 & Wells 1989).

As a general overview of the extent of forest across different types of environment in Tasmania, a cross-classification between the land systems and the eucalypt forest area data is presented in Table 2.7. This comparison reflects the generalised scale at which the published reports accurately correspond. It provides a means for presenting an indication of the extent of forest, relative to the land areas and land system diversity within classes of geology and altitude. However, it should be noted that the objective for assessing sampling adequacy is based on the finer-scale classification of eucalypt forest by a classification of geology and altitude within biogeographic regions (refer methods Section 2.2.3.2). Of the 462 land systems in Tasmania, 122 represent different categories of geology and altitude across seven biogeographic regions. Eucalypt-dominated forest vegetation is known from 76 of these geology-altitude categories, comprising 336 land systems. The most extensive areas of eucalypt forest in Tasmania are

associated with Jurassic dolerite rock types, which are also the most extensive land systems (Table 2.7). Eucalypt forests are largely predominant across lowland environments, but within substrate groupings, eucalypt forest may cover a greater proportion of the landscape extent in the mid-altitude range than in the lowland range (e.g. Jurassic dolerite, Mathinna beds, Permo-Triassic sediments, Quaternary deposits, and Quaternary till & talus).

Table 2.7 Matching between land system and eucalypt forest classifications for areas representing *ecoregions*. Relative extent of eucalypt forest [%Forest, Forest (ha); after Wells (1989) and Williams (1989)] compared with land area of geology and altitude classes [Land (ha); e.g. after Pemberton 1986]. The number of different land systems [*n*(LS)] and biophysical samples of eucalypt occurrence [*n*(plots)] are also indicated.

Geology group	Altitude class	Land (ha)	Forest (ha)	%Forest	*n*(LS)	*n*(plots)
Cambrian sediments	Sea-level to 600 m	144 572	64 675	45	10	408
Cambrian volcanics	Sea-level to 600 m	101 072	26 965	27	5	288
	600 m to 900 m	736	10	1	1	5
Devonian granite	Sea-level to 600 m	307 090	181 610	59	18	1 398
	600 m to 900 m	72 695	24 355	34	3	303
	900 m to tree-line	2 490	210	8	1	10
Jurassic dolerite	Sea-level to 600 m	1062 971	714 230	67	35	4 185
	600 m to 900 m	331 266	268 925	81	26	1 444
	900 m to tree-line	207 669	93 540	45	14	213
Mathinna beds	Sea-level to 600 m	220 255	140 852	64	12	1 391
	600 m to 900 m	9 145	7 460	82	1	107
Ordovician limestone	Sea-level to 600 m	29 793	19 825	67	4	111
	600 m to 900 m	1 062	30	3	1	0
Ordovician quartzites	Sea-level to 600 m	52 833	29 415	56	6	156
	600 m to 900 m	65 147	23 320	36	4	139
	900 m to tree-line	16 496	3 090	19	1	21
Permo-Triassic sediments	Sea-level to 600 m	596 898	257 590	43	53	1 950
	600 m to 900 m	135 810	66 775	49	11	195
	900 m to tree-line	11 980	360	3	1	0
Precambrian metamorphics	Sea-level to 600 m	239 764	68 445	29	5	53
	600 m to 900 m	77 667	19 330	25	3	112
	900 m to tree-line	25 311	975	4	2	6
Precambrian sediments	Sea-level to 600 m	412 291	163 725	40	16	1 093
Quaternary deposits	Sea-level to 600 m	665 937	120 102	18	56	527
	600 m to 900 m	1 785	1 245	70	1	1
Quaternary till & talus	600 m to 900 m	56 096	34 370	61	2	95
	900 m to tree-line	32 720	23 280	71	2	93
Tertiary basalt	Sea-level to 600 m	251 209	27 820	11	17	136
	600 m to 900 m	160 149	49 660	31	5	135
	900 m to tree-line	8 024	520	6	1	0
Tertiary sediments	Sea-level to 600 m	204 698	61 215	30	19	406

2.3.2 Characterising regional environmental heterogeneity

Regional environmental heterogeneity in Tasmania was characterised by the average cumulative relationship between land system diversity and area. Parameter estimates for the fit of the extreme value function (EVF) indicated that different non-linear relationships exist for five of the eight land system regions (Table 2.8). The King and Flinders Island regions (Richley 1984; Pinkard & Richley 1984) completely overlap in their 95% confidence intervals and could have been combined as a single region for the purpose of this analysis. All other regions were distinct for cumulative areas greater than about 55 000 ha, except for the eastern regions (Pinkard 1980, Davies 1988a) which overlap below about 220 000 ha.

Table 2.8 Regional models of environmental heterogeneity defined from the land systems. The relationships are defined by the Gompertz form of the extreme value function (see Box 2.1). E is the predicted number of different land systems, P is the maximum number of different land systems (asymptote), A is the land area, and the parameters β and γ are estimated.

Model Description		Parameters for $E = P\{1 - \exp(-\exp(\beta - \gamma \log_e A))\}$	
Land System Region	P	$-\beta$	$-\gamma$
King Island	16	4.590	0.524
Flinders Island	18	4.845	0.558
North West	93	6.496	0.639
North East	110	6.694	0.649
Central Plateau	55	6.993	0.750
South, East & Midlands	130	6.567	0.608
South West	40	5.589	0.583

2.3.3 Characterising local scale environmental heterogeneity

The local scale classification of biotic and abiotic attributes (Table 2.3) defined many more combinations of environment than actually existed, or which have been sampled. For the complete set of biophysical samples (n = 15 640), the local classification of physical attributes resulted in 2560 different types of abiotic environment (35.5% of possible types), and the classification of biotic attributes resulted in 1384 local habitats (7.6% of possible types), even though the potential for biotic variability was much greater than that for abiotic variability (refer Table 2.3). In the case of biotic environments, in particular, many of the 24 *Eucalyptus* species which have been recorded as forest dominants exist as such only under a very limited set of habitat conditions (e.g. *E. risdonii*, see Wiltshire *et al.* 1989). This is partly a reflection of the specialisation of some species for particular types of physical environment that further limits the biotic variability of their habitats.

The number of different local scale environments, resulting from separate classifications of either biotic or abiotic attributes of biophysical observations, was demonstrated for four eucalypt forest habitat units (Table 2.5). These habitat units represent landscape areas of different extent and patchiness characterised by different combinations of parent rock type, altitude and biogeographic region. On average, there were three or four biophysical observations of forest habitat for each type of abiotic environment, and five or six samples per class of biotic environment. In general, the number of different types of abiotic environment increased as the area of the habitat unit increased, but such a simple relationship was not apparent for the biotic environments.

The extrapolation to an asymptote for the cumulative relationship between abiotic or biotic heterogeneity with sample size provided an indication of *ecoseries* scale ecological variability and a basis for assessing local scale sampling adequacy (Table 2.9). The expected asymptote (P in Table 2.9) was always greater than the observed number of different types of environment (Abiotic or Biotic in Table 2.5) for the set of samples in each habitat unit.

Table 2.9 Ecoseries models of biotic and abiotic environmental heterogeneity defined from the biophysical samples in four eucalypt-forest habitat units. The relationships were defined by the Gompertz form of the extreme value function (see Box 2.1). E is the number of different biotic or abiotic environments, P is the iteratively-determined asymptote for the maximum number of different environments, A is the sampling frequency, and the parameters, β and γ, were estimated.

Model Description	Parameters for $E = P\{1\text{-}\exp(-\exp(\beta-\gamma\log_e A))\}$					
Habitat unit	Abiotic heterogeneity			Biotic heterogeneity		
Region Geology	P	β	γ	P	β	γ
South						
Permo-Triassic sediments	515	-5.659	-0.809	450	-5.537	-0.696
Jurassic dolerite	620	-5.535	-0.764	600	-5.580	-0.716
North-East						
Mathinna beds	550	-5.578	-0.780	550	-5.426	-0.728
Devonian granite	650	-5.688	-0.801	430	-5.114	-0.730

For the comparison of ecological processes, a straight line the simplest way of representing the relationship between biotic and abiotic environmental heterogeneity in each habitat unit. The slope of the line demonstrates differences in relative accumulation rates of biotic versus abiotic heterogeneity between habitat units (Table 2.10). These differences demonstrate local and regional variation in ecological processes due to the relative importance of substrate, climate, disturbance and biotic factors between sites. These spatio-temporal factors are in addition to attributes included in the existing classifications. For example, although the overall classification resulted in about half (54%) as many biotic as abiotic environments (i.e. 1384 biotic vs. 2560 abiotic for 15 640 samples), the rate of accumulation of biotic to abiotic environments was about 87% on Mathinna sediments in the North-East, 75% on Jurassic dolerite substrates in the South, and 70% on Devonian granites also in the North-East. Among these examples, only the ratio of biotic to abiotic environmental heterogeneity on Permo-Triassic sediments in the South (55%) approximated the overall rates of accumulation for the ecological dataset.

Table 2.10 Relationships between biotic and abiotic environmental heterogeneity in each of four eucalypt-forest habitat units.

Parameters and model fit statistics for the linear regression: Biotic = $\beta \times$ *Abiotic*		
Habitat unit	β	R^2_{adj}
South		
Permo-Triassic sediments	0.5516	0.9978
Jurassic dolerite	0.7474	0.9989
North-East		
Mathinna beds	0.8735	0.9991
Devonian granite	0.6970	0.9979

These results indicated that in three of the four cases demonstrated, sampling in proportion to a classification of abiotic environmental heterogeneity, may not adequately reflect the spatio-temporal processes influencing ecological variability in all habitat types. Other processes influence the structural and floristic diversity of vegetation. For example, the diversity of vegetation on Mathinna sediments in the North-East may be, in part, due to the grouping of argillaceous and arenaceous sediments across a variety of land-forms from lowland to subalpine

habitats, and in part due to different fire regimes and a variety of other disturbances linked to human land use. Therefore, for the purposes of assessing sampling adequacy, it is important that both biotic and abiotic factors be used to define a classification of ecological variability.

2.3.4 Assessing sampling adequacy in eucalypt forest habitat units

2.3.4.1 Expected sampling frequency: interpolating regional environmental heterogeneity

The regional models of environmental heterogeneity, defined from the land systems classification for Tasmania, in combination with the rules for unexplained spatio-temporal variability (Table 2.4), resulted in the interpolation of 7974 different types of habitat for the 110 forest areas (Table 2.11), summarised in Table 2.12. The seven land system regions were matched to the comparable biogeographic region in which eucalypt forest habitat units were classified. The two northern regions (Northwest and North & North Midlands) being approximated by the model of environmental heterogeneity for the North-West land systems region.

Table 2.11 Sampling adequacy of 110 eucalypt forest habitat units across eight regions in Tasmania. Estimates of sampling requirements ($\hat{n}$) are calculated as one replicate (two samples) per predicted environmental heterogeneity in each habitat unit, stratified by spatial extent and adjusted for over-prediction in small areas. The deviance statistic with the Chi-squared test is used to assess significant differences. Sampling within habitat units within regions is under represented if standardised residuals, $r_s < -2$ (denoted by the symbol, '-'), or over-represented if $r_s > +2$ (denoted by the symbol, '+'), otherwise the symbol '*' is used for $0.05 > p > 0.1$; and 'ns' for $p > 0.1$. The number of additional samples required to achieve minimum expected replication and representation [n(add)] is indicated for each habitat unit.

Region, Geology & Altitude categories		Area (ha)	E_n	n	$\hat{n}$	r_s	p	n(add)
South								
Cambrian sediments	Sealevel to 600 m	23785	62	7	124	-10 5	-	117
Jurassic dolerite	Sealevel to 600 m	115390	106	1464	848	21.2	+	0
Jurassic dolerite	600 m to 900 m	36850	74	300	296	0.2	ns	0
Jurassic dolerite	900 m to treeline	2990	22	7	43	-5.5	-	36
Ordovician limestone	Sealevel to 600 m	11135	43	60	87	-2.9	-	27
Ordovician quartzites	Sealevel to 600 m	1280	13	0	27	-5.2	-	27
Ordovician quartzites	600 m to 900 m	5670	31	8	62	-6.8	-	54
Permo-Triassic sediments	Sealevel to 600 m	72375	94	929	561	15.5	+	0
Permo-Triassic sediments	600 m to 900 m	2950	22	44	43	0.1	ns	0
Permo-Triassic sediments	900 m to treeline	360	6	0	13	-3.6	-	13
Precambrian metamorphics	Sealevel to 600 m	20605	58	16	116	-9.3	-	100
Precambrian sediments	Sealevel to 600 m	29815	68	245	272	-1.7	*	27
Quaternary deposits	Sealevel to 600 m	14115	49	35	98	-6.3	-	63
Quaternary deposits	600 m to 900 m	1245	13	1	26	-5	-	25
Tertiary sediments	Sealevel to 600 m	2195	18	0	37	-6	-	37
Deviance = 2207, df = 14, p << 0.005		*340760*	*679*	*3116*	*2653*			*526*
West								
Cambrian sediments	Sealevel to 600 m	11400	41	39	83	-4.8	-	44
Cambrian volcanics	Sealevel to 600 m	7400	34	8	67	-7.2	-	59
Devonian granite	Sealevel to 600 m	2710	20	2	39	-5.9	-	37
Jurassic dolerite	Sealevel to 600 m	1570	14	0	29	-5.3	-	29
Ordovician limestone	Sealevel to 600 m	1180	12	4	24	-4 1	-	20
Ordovician quartzites	Sealevel to 600 m	12580	43	6	87	-8 7	-	81
Ordovician quartzites	600 m to 900 m	1635	15	2	29	-5	-	27
Permo-Triassic sediments	Sealevel to 600 m	2135	17	15	34	-3.3	-	19
Permo-Triassic sediments	600 m to 900 m	755	9	0	18	-4.3	-	18
Precambrian metamorphics	Sealevel to 600 m	37800	67	0	268	-16.4	-	268
Precambrian metamorphics	600 m to 900 m	5895	30	0	60	-7.7	-	60
Precambrian sediments	Sealevel to 600 m	65765	78	49	467	-19.3	-	418
Quaternary deposits	Sealevel to 600 m	10865	41	11	81	-7.8	-	70
Tertiary basalt	Sealevel to 600 m	1515	14	3	28	-4.7	-	25
Tertiary sediments	Sealevel to 600 m	5120	28	0	55	-7 4	-	55
Deviance = -465, df = 14, p << 0 005		*168325*	*463*	*139*	*1369*			*1230*
Centre								
Cambrian sediments	Sealevel to 600 m	340	4	8	8	0.1	ns	0
Cambrian volcanics	Sealevel to 600 m	1620	11	13	23	-2.1	-	10
Devonian granite	600 m to 900 m	420	4	0	9	-3	-	9
Jurassic dolerite	Sealevel to 600 m	58120	53	112	319	-11.6	-	207
Jurassic dolerite	600 m to 900 m	135560	55	613	439	8.3	+	0
Jurassic dolerite	900 m to treeline	84605	54	149	327	-9.8	-	178
Ordovician limestone	Sealevel to 600 m	2040	13	2	27	-4.8	-	25
Ordovician quartzites	Sealevel to 600 m	230	3	0	6	-2.4	-	6
Ordovician quartzites	600 m to 900 m	10015	33	42	66	-3	-	24
Ordovician quartzites	900 m to treeline	2710	16	19	32	-2.3	-	13
Permo-Triassic sediments	Sealevel to 600 m	26965	47	73	188	-8.4	-	115
Permo-Triassic sediments	600 m to 900 m	48345	52	70	209	-9 6	-	139
Precambrian metamorphics	Sealevel to 600 m	1730	12	0	24	-4 9	-	24
Precambrian metamorphics	600 m to 900 m	10985	34	53	69	-1.9	*	16
Precambrian metamorphics	900 m to treeline	975	8	6	16	-2.5	-	10
Precambrian sediments	Sealevel to 600 m	1470	11	0	22	-4.6	-	22
Quaternary deposits	Sealevel to 600 m	2090	14	8	27	-3.7	-	19
Quaternary till & talus	600 m to 900 m	34370	50	95	198	-7.3	-	103
Quaternary till & talus	900 m to treeline	23020	45	92	90	0.2	ns	0
Tertiary basalt	Sealevel to 600 m	1385	10	6	21	-3.2	-	15
Tertiary basalt	600 m to 900 m	49090	52	125	209	-5.8	-	84
Tertiary basalt	900 m to treeline	520	5	0	10	-3.2	-	10
Deviance = -773, df = 21; p << 0.005		*496605*	*589*	*1486*	*2340*			*1029*

Table 2.11 (continued from previous page). Sampling adequacy of 110 eucalypt forest habitat units across eight regions in Tasmania.

Region, Geology & Altitude categories		Area (ha)	E_n	n	$\bar{n}$	r_s	p	n(add)
Northwest								
Cambrian sediments	Sealevel to 600 m	28260	61	354	243	7.2	+	0
Cambrian volcanics	Sealevel to 600 m	17035	50	254	99	15.5	+	0
Devonian granite	Sealevel to 600 m	11020	41	102	82	2.3	+	0
Jurassic dolerite	Sealevel to 600 m	2995	21	36	41	-0.8	ns	0
Ordovician limestone	Sealevel to 600 m	2610	19	5	38	-5 4	-	33
Ordovician quartzites	Sealevel to 600 m	10115	39	115	78	4.1	+	0
Ordovician quartzites	600 m to 900 m	3890	24	67	48	2.8	+	0
Ordovician quartzites	900 m to treeline	380	6	2	12	-2.9	-	10
Permo-Triassic sediments	Sealevel to 600 m	17220	50	184	100	8.4	+	0
Precambrian metamorphics	Sealevel to 600 m	6820	32	29	64	-4.4	-	35
Precambrian metamorphics	600 m to 900 m	1800	15	43	31	2.2	+	0
Precambrian sediments	Sealevel to 600 m	51590	73	555	440	5.5	+	0
Quaternary deposits	Sealevel to 600 m	19365	52	86	105	-1.8	*	19
Tertiary basalt	Sealevel to 600 m	19025	52	99	104	-0.5	ns	0
Tertiary basalt	600 m to 900 m	390	6	10	12	-0.7	ns	0
Tertiary sediments	Sealevel to 600 m	590	8	0	16	-4	-	16
Deviance = 1304; df =15 ; p << 0.005		*193105*	*549*	*1941*	*1513*			*125*
North & Northern Midlands								
Cambrian sediments	Sealevel to 600 m	890	11	0	21	-4.6	-	21
Cambrian volcanics	Sealevel to 600 m	920	11	17	21	-1	ns	4
Jurassic dolerite	Sealevel to 600 m	84540	94	497	561	-2.7	-	64
Jurassic dolerite	600 m to 900 m	7380	36	167	72	11.2	+	0
Ordovician limestone	Sealevel to 600 m	2890	21	40	43	-0.4	ns	3
Ordovician quartzites	Sealevel to 600 m	5210	30	35	60	-3.2	-	25
Ordovician quartzites	600 m to 900 m	2110	18	20	36	-2.6	-	16
Permo-Triassic sediments	Sealevel to 600 m	25465	64	221	258	-2.3	-	37
Permo-Triassic sediments	600 m to 900 m	6670	34	23	68	-5.5	-	45
Precambrian metamorphics	Sealevel to 600 m	1490	14	8	29	-3.9	-	21
Precambrian metamorphics	600 m to 900 m	650	9	16	17	-0.3	ns	1
Precambrian sediments	Sealevel to 600 m	15085	51	244	103	13.9	+	0
Quaternary deposits	Sealevel to 600 m	17210	55	57	109	-5	-	52
Quaternary till & talus	900 m to treeline	260	5	1	10	-2.8	-	9
Tertiary basalt	Sealevel to 600 m	1420	14	7	28	-4	-	21
Tertiary sediments	Sealevel to 600 m	16070	53	59	106	-4.5	-	47
Deviance = 201; df =15 , p << 0.005		*188260*	*520*	*1412*	*1542*			*366*
East & Southern Midlands								
Devonian granite	Sealevel to 600 m	18745	56	57	111	-5.2	-	54
Jurassic dolerite	Sealevel to 600 m	382410	126	1892	1261	17 8	+	0
Jurassic dolerite	600 m to 900 m	57505	87	182	522	-14.9	-	340
Mathinna beds	Sealevel to 600 m	9495	40	7	80	-8.2	-	73
Permo-Triassic sediments	Sealevel to 600 m	104540	104	503	828	-11.3	-	325
Quaternary deposits	Sealevel to 600 m	11520	44	116	88	2.9	+	0
Tertiary basalt	Sealevel to 600 m	2215	18	14	37	-3.8	-	23
Tertiary sediments	Sealevel to 600 m	680	9	6	19	-2 9	-	13
Deviance = 563, df =7 ; p << 0.005		*587110*	*484*	*2777*	*2946*			*828*
Northeast								
Devonian granite	Sealevel to 600 m	140875	102	1196	814	13.4	+	0
Devonian granite	600 m to 900 m	23935	63	303	126	15.8	+	0
Devonian granite	900 m to treeline	210	4	10	9	0 5	ns	0
Jurassic dolerite	Sealevel to 600 m	69205	89	184	536	-15.2	-	352
Jurassic dolerite	600 m to 900 m	31630	70	182	280	-5.9	-	98
Jurassic dolerite	900 m to treeline	5945	32	57	64	-0.9	ns	7
Mathinna beds	Sealevel to 600 m	126645	100	1367	803	19.9	+	0
Mathinna beds	600 m to 900 m	7460	36	107	72	4 1	+	0
Permo-Triassic sediments	Sealevel to 600 m	8890	40	25	79	-6 1	-	54
Permo-Triassic sediments	600 m to 900 m	8055	38	58	75	-2	-	17
Quaternary deposits	Sealevel to 600 m	39995	76	173	305	-7.5	-	132
Tertiary basalt	Sealevel to 600 m	2260	18	7	37	-4.9	-	30
Tertiary basalt	600 m to 900 m	180	4	0	8	-2.8	-	8
Tertiary sediments	Sealevel to 600 m	36560	74	341	295	2.6	+	0
Deviance = 2220; df = 13, p << 0.005		*501845*	*746*	*4010*	*3503*			*691*
King Island & Furneaux Islands								
Devonian granite	Sealevel to 600 m	8260	13	41	25	3.1	+	0
Mathinna beds	Sealevel to 600 m	4712	11	17	21	-0.9	ns	4
Quaternary deposits	Sealevel to 600 m	4942	11	41	21	4.2	+	0
Precambrian sediments	Sealevel to 600 m	500	4	0	8	-2 8	-	8
Deviance = 86, df = 3; p << 0 005		*18414*	*39*	*99*	*7567*			*8*

The existing set of biophysical samples contain 14 980 observations of eucalypt presence. On average, the observed sampling frequencies represent a replication rate of 2.3 times the predicted levels of environmental heterogeneity at the *ecosection* to *ecodistrict* scale (Fig. 2.4). The inclusion of one replicate (two samples) for each predicted level of environmental heterogeneity associated with eucalypt forest results in an expected sample size that approximates the observed sample size of the ecological date-set (i.e. about 15 000 biophysical observations). Therefore, predictions based on the existing set of ecological data can only be applied at the coarser scale of the *ecosection* to *ecodistrict* rather than the original objective of the *ecoseries* scale.

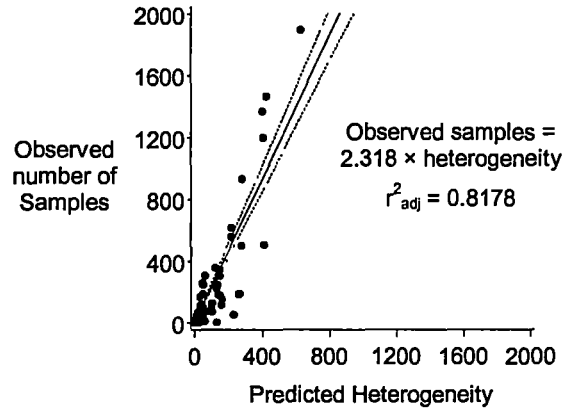


Figure 2.4 Average replication rate within the sample of observations, relative to predicted environmental heterogeneity. Predicted environmental heterogeneity is derived from regional models of land system accumulation.

Assuming biophysical observations were randomly located within eucalypt forest habitat units, a Chi-squared test of significance between observed and expected sampling frequencies provides an assessment of sampling adequacy at the coarser scale of *ecosections* to *ecodistricts* (Table 2.11). This assessment of sampling adequacy is displayed relative to the reference line of sampling equivalence in Fig. 2.5. The overall comparison of the distribution of expected and observed sampling frequencies between habitat units within a region was significantly different ($p << 0.005$) for all eight regions (Table 2.12). However within each region, the observed sampling frequency for some habitat units was consistent with expected sampling frequencies (Table 2.11).

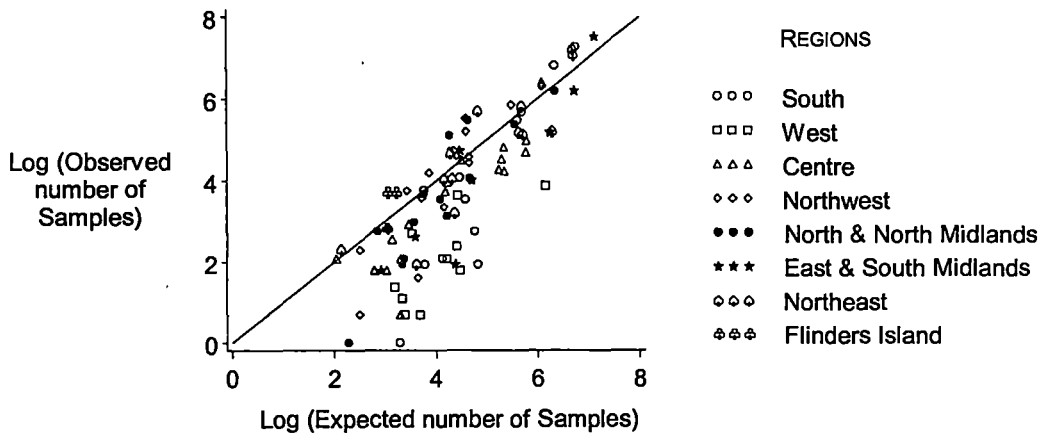


Figure 2.5 The natural log relationship between observed and expected sampling frequency for habitat units within each biogeographic region. The standard diagonal line defines equivalence between observed and expected sampling, above which sampling is more than expected, and below which sampling is less than expected.

Table 2.12 Summary of regional sampling adequacy of environmental heterogeneity within eucalypt forest habitat units at the *ecosection* to *ecodistrict* scale of ecosystem classification. A Chi-squared test of significance for consistency between observed (n) and minimum expected random sampling frequencies ($\hat{n}$) within individual habitat units indicated that in no region does observed sampling completely represent the expected levels of environmental heterogeneity ($p << 0.005$). The number of additional samples required to achieve minimum expected replication and representation is summarised across habitat units within regions.

Biogeographic Region	Forest (ha)	Habitat units	n	$\hat{n}$	Additions
Furneaux Islands	18 414	4	99	75	8
Northwest	193 105	16	1941	1513	125
North & North Midlands	188 260	16	1412	1542	366
Northeast	501 845	14	4010	3503	699
East & South Midlands	587 110	8	2777	2946	828
Centre	496 605	22	1486	2340	1029
West	168 325	15	139	1369	1230
South	340 760	15	3116	2653	526
Overall results	2 494 424	110	14 980	15 941	4811

Although the total number of samples expected for the eight regions within Tasmania were approximately comparable with the observed number of samples (i.e. 15 941 samples vs. 14 980 samples), an additional 4811 samples are required across the 110 habitat units to achieve minimum levels of replication (Table 2.12). This relatively high requirement for additional sampling is due to bias within the observed sample that comprises 3853 'redundant' biophysical observations (*cf.* Fig. 2.5). The frequency with which additional sampling would be randomly located within each habitat unit in each region was indicated in Table 2.11. If replication is increased to two (i.e. three samples for each predicted level of heterogeneity), then the expected sampling frequency increases to *c.* 23 900 samples. However, because relatively high levels of sampling exist in some habitat units, only 9780 additional samples would be required to achieve this level of representativeness. The corresponding 'redundancy' reduces to 857 samples.

Observations of eucalypt occurrence in the ecological dataset are present in 93 of the 110 habitat-units (Table 2.11). Of the 17 unsampled habitat-units, four represent areas that are less than 500 ha, but 13 appear to be significant omissions in the sampling of eucalypt occurrence considering their extent in the landscape (500 to 37 800 ha). In the 'South' region, the areas that represent such significant omissions are lowland (< 600 m) Ordovician quartzites and Tertiary sediments. In the 'West' these are habitat-units defined by areas of Jurassic dolerite, Precambrian metamorphics and Tertiary sediments, and mid-altitude range (600 to 900 m) Permo-Triassic sediments. In the 'North & North Midlands' these are Cambrian sediments, and in the 'Centre' these are Precambrian sediments, and lowland (≤ 600 m) Precambrian metamorphics.

Minimum levels of sampling are achieved for 38 of the 110 habitat units at the *ecosection* to *ecodistrict* scale (Table 2.11). The level of observed and expected sampling is consistent (> 0.05) for 16 of the 110 habitat units, and a further 22 habitat units are sampled in excess of this requirement. Of the remaining 72 habitat units, 12 are of relatively small extent and require fewer than 20 randomly located observations to minimally sample expected levels of environmental heterogeneity. At least 100 observations have been made in each of a further 10 habitat units (with up to 503 samples on lowland Permo-Triassic sediments from the 'East & Southern Midland' region), and there are at least 30 observations in a further 13 habitat units. Overall, 73 of the 110 habitat units could be considered sufficiently sampled to either comprehensively reflect the ecological variability or at least capture a reasonable level of the expected environmental heterogeneity associated with eucalypt forest occurrence. Excluding the four habitat units of minor extent (≤ 500 ha), this leaves 33 areas that are seriously under-sampled. Although some of these under-sampled habitat units ($r_S > -2$) have reasonably large samples ($n \geq 30$), they are unlikely to comprehensively represent eucalypt forest occurrence (Table 2.11). Therefore, unrepresented or poorly represented habitat units were defined as under-sampled areas ($r_S > -2$) of moderate to large extent (≥ 500 ha) with either fewer than 30 existing observations or requiring more than 100 additional observations to account for the expected levels of environmental heterogeneity (Table 2.13).

Questions of eucalypt ecology at the *ecosection* to *ecodistrict* scale that are based on correlative analyses of the existing set of data cannot be confidently addressed in 46 habitat units (Table 2.13). These habitat units largely represent the environments of western, central and south-western Tasmania. The only typically eastern habitat unit of concern for the study of eucalypt ecology is Mathinna beds in the East Coast and Tiers Nature Conservation Region (*cf.* NCR 7a, Williams 1989; 'East & Southern Midlands' region). However, this habitat type is juxtaposed with similar environments of the 'North East' region, and therefore predictions based on data from either region are likely to reasonably apply to the other.

Table 2.13 Unrepresented and poorly represented habitat units in the ecological dataset for the occurrence of eucalypt forest in Tasmania. Data summarised from Table 2.11. Habitat units in which ecological variability may be reasonably represented, despite the requirement for a large number of additional randomly located samples, are indicated by an asterisk.

Region, Geology & Altitude categories		Area (ha)	*n*	Additions
South				
Cambrian sediments	Sea-level to 600 m	23 785	7	117
Jurassic dolerite	900 m to tree-line	2 990	7	36
Ordovician quartzites	Sea-level to 600 m	1 280	0	27
Ordovician quartzites	600 m to 900 m	5 670	8	54
Precambrian metamorphics	Sea-level to 600 m	20 605	16	100
Quaternary deposits	600 m to 900 m	1 245	1	25
Tertiary sediments	Sea-level to 600 m	2 195	0	37
West				
Cambrian volcanics	Sea-level to 600 m	7 400	8	59
Devonian granite	Sea-level to 600 m	2 710	2	37
Jurassic dolerite	Sea-level to 600 m	1 570	0	29
Ordovician limestone	Sea-level to 600 m	1 180	4	20
Ordovician quartzites	Sea-level to 600 m	12 580	6	81
Ordovician quartzites	600 m to 900 m	1 635	2	27
Permo-Triassic sediments	Sea-level to 600 m	2 135	15	19
Precambrian metamorphics	Sea-level to 600 m	37 800	0	268
Precambrian metamorphics	600 m to 900 m	5 895	0	60
Precambrian sediments	Sea-level to 600 m	65 765	*49	418
Quaternary deposits	Sea-level to 600 m	10 865	11	70
Tertiary basalt	Sea-level to 600 m	1 515	3	25
Tertiary sediments	Sea-level to 600 m	5 120	0	55
Centre				
Cambrian volcanics	Sea-level to 600 m	1 620	13	10
Jurassic dolerite	Sea-level to 600 m	58 120	*112	207
Jurassic dolerite	900 m to tree-line	84 605	*149	178
Ordovician limestone	Sea-level to 600 m	2 040	2	25
Ordovician quartzites	900 m to tree-line	2 710	19	13
Permo-Triassic sediments	Sea-level to 600 m	26 965	*73	115
Permo-Triassic sediments	600 m to 900 m	48 345	*70	139
Precambrian metamorphics	Sea-level to 600 m	1 730	0	24
Precambrian sediments	Sea-level to 600 m	1 470	0	22
Quaternary deposits	Sea-level to 600 m	2 090	8	19
Quaternary till & talus	600 m to 900 m	34 370	*95	103
Tertiary basalt	Sea-level to 600 m	1 385	6	15
Northwest				
Ordovician limestone	Sea-level to 600 m	2 610	5	33
Precambrian metamorphics	Sea-level to 600 m	6 820	29	35
North & Northern Midlands				
Cambrian sediments	Sea-level to 600 m	890	0	21
Ordovician quartzites	600 m to 900 m	2 110	20	16
Permo-Triassic sediments	600 m to 900 m	6 670	23	45
Precambrian metamorphics	Sea-level to 600 m	1 490	8	21
Tertiary basalt	Sea-level to 600 m	1 420	7	21
East & Southern Midlands				
Jurassic dolerite	600 m to 900 m	57 505	*182	340
Mathinna beds	Sea-level to 600 m	9 495	7	73
Permo-Triassic sediments	Sea-level to 600 m	104 540	*503	325
Tertiary basalt	Sea-level to 600 m	2 215	14	23
Northeast				
Tertiary basalt	Sea-level to 600 m	2 260	7	30
Jurassic dolerite	Sea-level to 600 m	69 205	*184	52
Quaternary deposits	Sea-level to 600 m	39 995	*173	132

2.3.4.2 Expected sampling frequency: extrapolating local environmental heterogeneity

Comparison between observed and predicted levels of abiotic and biotic heterogeneity indicated that existing sampling in each of four habitat units may represent only about 38% to 74% of the potential ecoseries scale ecological variability (Table 2.14). Representation of biotic environments in the ecological dataset for the two habitat units from the South appears to be

poorest with 38% and 56% of potential variation in the vegetation being present for each set of observations.

Table 2.14 Representation of local scale biotic and abiotic asymptotic heterogeneity within four habitat units (contrast *P* in Table 2.8 with *n*(abiotic) or *n*(biotic) in Table 2.4).

Habitat unit	Percentage Representation	
Region Geology	Biotic Heterogeneity	Abiotic Heterogeneity
South		
Permo-Triassic sediments	38.2	60.2
Jurassic dolerite	55.7	70.6
North-East		
Mathinna beds	60.0	68.4
Devonian granite	74.0	72.2

Extrapolation of the characteristic curves for local scale biotic and abiotic environmental heterogeneity (defined in Table 2.9) enabled the number of samples required to represent particular levels of environmental heterogeneity to be estimated. Assuming random sampling, scenarios based upon representation of 75%, 95% or 99% (shown as $\hat{n}_{E_{75}}$, $\hat{n}_{E_{95}}$ and $\hat{n}_{E_{99}}$ in Table 2.15) of either biotic or abiotic heterogeneity, at the *ecoseries* scale, are indicated in Table 2.15. A sampling frequency that approximates 75% of the abiotic heterogeneity is most similar to the current sampling levels for the four habitat units within the ecological dataset (i.e. compare *n*(sample) in Table 2.5 with $\hat{n}_{E_{75}}$ in Table 2.15). At the local scale, 75% of potential levels of heterogeneity is a reasonable objective of random sampling. However, the observed sampling frequencies of biotic and abiotic heterogeneity in each habitat unit are usually much less than the expected sample size for obtaining 75% of the predicted environmental heterogeneity (compare to Table 2.14).

Table 2.15 Scenarios for the expected number of random samples (to the nearest 10 observations) that would capture 75%, 95% or 99% of the abiotic or biotic heterogeneity estimated by the asymptote of the EVF models for *ecotope* to *ecoseries* scale ecological variability from the existing sample of observations in four habitat units (see Table 2.4). Note that $\hat{n}_{E_{75}}$, $\hat{n}_{E_{95}}$ and $\hat{n}_{E_{99}}$ indicate the number of samples required to represent 75%, 95% and 99% respectively of the abiotic or biotic heterogeneity

Habitat unit	Biotic Heterogeneity			Abiotic Heterogeneity		
Region Geology	$\hat{n}_{E_{75}}$	$\hat{n}_{E_{95}}$	$\hat{n}_{E_{99}}$	$\hat{n}_{E_{75}}$	$\hat{n}_{E_{95}}$	$\hat{n}_{E_{99}}$
South						
Permo-Triassic sediments	4 550	13 760	25 510	1 650	4 260	7 250
Jurassic dolerite	3 840	11 260	20 530	2 150	5 890	10 340
North-East						
Mathinna beds	2 700	7 780	14 040	1 940	5 200	9 020
Devonian granite	1 730	4 960	8 940	1 830	4 770	8 150

The four habitat units vary in the comparison of observed and expected sampling frequencies based on the objective of 75% of the potential abiotic or biotic environmental heterogeneity

(Table 2.16). In general, the existing set of samples comprises 60% to 90% of abiotic and 45% to 95% biotic heterogeneity. At the extremes of this comparison, the observed set of observations on Permo-Triassic sediments in the South represent only 22% of the sample required to comprise 75% of the asymptotic biotic heterogeneity, and on Devonian granite in the Northeast, the existing set of observations represent nearly 95% of the sample required for 75% of the biotic heterogeneity. Based on random sampling, these results indicate that even the most frequently sampled habitat units within the ecological dataset are incompletely represented for even 75% of *ecoseries* level biotic and abiotic environmental heterogeneity.

Table 2.16 Local scale sampling adequacy for representation of 75% of asymptotic local scale biotic and abiotic heterogeneity within four habitat units (compare Table 2.5 with Table 2.15). The additional number of randomly-located samples required to represent 75% of potential biotic or abiotic heterogeneity is also shown.

Habitat unit	Biotic Heterogeneity		Abiotic Heterogeneity	
Region Geology	% Represented	Additions	% Represented	Additions
South				
Permo-Triassic sediments	21.6	3565	59.7	665
Jurassic dolerite	47.0	2032	84.1	342
North-East				
Mathinna beds	55.7	1195	77.6	435
Devonian granite	94.5	95	89.3	195

2.4 Discussion

An assessment of the sampling adequacy of ecological data, whether resulting from an *ad hoc* compilation, or derived from a designed survey, should be a prerequisite of predictive modelling. This is especially important where predictions depend upon the frequency of different vegetation-by-environment interactions to define the patterns of a response (e.g. Austin & Meyers 1996). Data representativeness can be comprehensively assessed, for example, by reference to the sampling distributions expected from an ideal survey design, in which the scale and attributes used in a stratification of landscapes are consistent with the experimental question and the purposes for which the predictions are made. Assuming that a sampling distribution which is proportional to the landscape levels of heterogeneity adequately represents the relationship between vegetation and environment, then an assessment of sampling adequacy reduces to an assessment of ecological variability. However, an assessment of ecological variability will frequently be limited to the scale and attributes for which landscape information has been comprehensively defined (e.g. see articles in Sims 1996 and Sims *et al.* 1996).

In this analysis of ecological representativeness in a set of compiled data for eucalypt forest in Tasmania, the objective was to summarise ecological variability for local land areas that represented processes at the *ecoseries* scale. However, two existing information sources were only available for coarser scales, *ecosection* to *ecodistrict*, but were unmatched at this classification scale. Given this limitation, a statistical method was derived to enabled the most

important components of both classifications to be utilised in an assessment of expected regional levels of ecological variability. The classification of eucalypt forest habitats by generalised groupings of geology and altitude (Wells 1989; Williams 1989) defined the primary network of regional land units with specified areas. The land systems, which comprised more detail about the variability and extent of physical environments in each region were used to define relationship between land classification diversity and land area. The average of these relationships were found to be well-approximated by an extreme value function (EVF), that was recommended for use by Williams (1995, 1996) in cases related to species-area type questions. The application of species-area theory to an analysis of ecosystem heterogeneity is consistent with the accumulating evidence for the relationship between biological variability and spatio-temporal variability of the associated environment (e.g. Williamson 1988, 1989; Williamson & Lawton 1991).

The resulting estimates of environmental heterogeneity within each eucalypt forest habitat unit were statistically indicative of the potential habitat diversity, but the actual attributes defining each type of environment were unknown. However, the land systems represent subjectively drawn map boundaries that were derived by consensus with unclear choice and weighting of source information. These classifications have variable accuracy and may not be reliable for use in different circumstances. The fixed boundary classifications of the land systems (e.g. Pemberton 1986) and the corresponding biogeographic regions (e.g. Orchard 1988; Thackway & Cresswell 1995) are useful summaries of the complex patterns of the spatial concordance among ecosystems and environments. However, in reality, these land units are relatively arbitrary entities and continuous variation exists across boundaries with transitions in some cases being abrupt, and in others cases being broad or diffuse, such that the placement of boundaries is a subjective matter of judgment (Williams, P. 1996).

Greater flexibility and repeatability of combining and recombining different landscape features is needed to more accurately assess sampling adequacy at the scale and for the attributes that are of interest to the analysis. This is currently possible with computer systems, given the appropriate scale and attributes of environmental information for the particular analysis (e.g. Cocks & Baird 1991; Michaelson *et al.* 1994; Neldner *et al.* 1995; Host *et al.* 1996). More detailed approaches to assessing the ecological variability of a landscape are also feasible with higher-resolution models of climate (e.g. Glassy & Running 1994; Airey & Hulme 1995; Fanning & Weaver 1996; Guenni *et al.* 1996; Mackey *et al.* 1996; Guenni 1997) and terrain processes (e.g. Skidmore 1989; Dubayah 1994; Merot *et al* 1995; Mitasova *et al.* 1996). However, these techniques require specialised computing support for large databases linked to geographic information systems (Margules *et al.* 1994; see also discussion by McDonnell 1996; and applications by Moore *et al.* 1993; Adinarayana & Krishna 1995; Host *et al.* 1996). Therefore, for the relatively simple and routine purpose of assessing sampling adequacy, prior to ecological analyses, it may be more

prudent to adopt an existing land classification and develop surrogate variables and apply statistical methods to account for inadequacies in the scale or attributes used.

Species-area theory appeared applicable to this assessment of sampling adequacy at contrasting scales. In addition to the interpolation of heterogeneity between regional information sources with common underlying ecological processes, rarefaction was used to extrapolate sampled levels of either biotic or abiotic heterogeneity from a classification of the respective attributes representing an *ecoseries* scale. This rarefaction analysis was more comparable with the original scale at which prediction and mapping of vegetation patterns was intended (e.g. 1:25 000 to 1:100 000 map scales). In fact, the existing sample of ecological data for eucalypt occurrences was found to represent a classification scale between *ecosection* and *ecodistrict* rather than *ecoseries*. Thus, the assessment of sampling adequacy was also able to determine not only those spatial regions in which sampling was representative, but also the scale at which subsequent predictive analyses might be expected to be reasonably precise. This is similar to undertaking a homogeneity analysis when defining landscape classification scales for maps (e.g. Bedward *et al.* 1992), or in accuracy assessments for remote sensing (e.g. Stehman 1996).

The application of rarefaction type analyses in this chapter enabled increasing environmental heterogeneity to be equated with increasing spatial scale, being either area or number of pooled samples, and to either set the asymptote or estimate it as a model parameter. The extreme value function (EVF) (after Williams 1995, 1996) was thus equally applicable to a description of the spatial variation in biophysical diversity at contrasting scales. Therefore, for a particular classification of environmental heterogeneity, the differences in asymptote and other parameter estimates that characterise the shape of the accumulation curves in different biogeographic regions, or in different local areas of habitat within regions, is indicative of different ecological processes (e.g. Bullock *et al.* 1995). The different EVF relationships for biotic and abiotic heterogeneity within eucalypt forest habitat units, for instance, indicated the importance of combining both these types of attributes when estimating ecological variability. In addition, the different linear relationships between biotic and abiotic heterogeneity curves between habitat units were indicative of probable local scale differences in ecological processes within and between biogeographic regions.

The coincidence of the species-area type curve in many situations suggests a common scaling function in nature. The search for fundamental scaling rules in nature is a general objective of assessments of spatial heterogeneity (e.g. Li & Reynolds 1994; Pacala & Deutschman 1995) or fractal studies of landscape structure (e.g. Burrough 1981; Scheuring & Riedi 1994; Loehle & Wein 1994; Gao & Xia 1996). However, further work is needed to more comprehensively explore the potential of environmental accumulation curves in survey design and their general application to assessments of sampling adequacy, providing a link between different ecosystem scales and ecological classifications.

The results for the assessment of sampling adequacy for eucalypt forest habitats in Tasmania indicated that for the *ecosection* to *ecodistrict* scale of classification (Table 2.1), about 81% of habitat units comprised some representation of expected levels of ecological variability, but these were largely the lowland habitats and eastern regions (see section 2.3.4.1). Eucalypt forest habitats in western, central and highland regions were largely unsampled or undersampled. Therefore, considering the occurrence of eucalypt forest in Tasmania as a whole, the current sample suggests a high degree of bias. Such bias does not preclude predictive modelling, but it should not be ignored (Austin & Meyers 1996). For example, questions of eucalypt forest ecology could be restricted to the more representative regions of eastern and lowland habitats. On a spatial scale, these are extensive areas in Tasmania.

Some of the bias inherent to this sample reflects the sampling objectives of Forestry Tasmania in which a spatial stratification of the landscape based on forest type characteristics was devised for their purposes of representative sampling among the production forests. This is the basis objective of a Continuous Forest Inventory (Lawrence 1978; see also Scott & Kohl 1994). The original forest type stratification aimed to ensure a representative sample of forests of wood-production potential, as a bioassay of variation in environmental conditions. However, the current purpose for analysis of these data is to explore questions of eucalypt ecology from the inherent relationships between eucalypt forest occurrence and environment. In this respect, the physical parameters that limit vegetation occurrence are as much of interest as the similarity of response to different environmental conditions, since these characterise the correlation between the species' response and habitat type. Although this ideal has not been completely achieved, the quantitative assessment of sampling adequacy has enabled specific options for dealing with sampling bias to be considered.

There are two main options for dealing with sampling bias, either explore possibilities for additional sampling or explicitly set constraints on statistical analysis and caveats on the interpretation of subsequent predictions.

If predictions are intended for the *ecoseries* scale, then additional sampling is needed. The current assessment assumed that sampling within eucalypt-forest habitat units was completely random. This is because the characteristic environments within these areas were unknown, largely because of the statistical approach to analysis. However, the appropriate land information would enable the environments within habitat units to be clearly identified and stratified, within which samples could be more efficiently located in the landscape, while maintaining randomness within classification strata. A suitable field survey design for supplementary sampling would the GRADSECT procedure (e.g. Margules & Austin 1994), since efficiency in locating samples will be important. The resulting set of samples could be checked for the success of the field survey objectives, following the type of approach considered here, given that bias was necessarily introduced for the purpose of cost-effective sampling (Austin & Heyligers 1991). Species-area approaches have been applied by other workers to monitor the progress toward representative

sampling of a field survey (e.g. accumulation of vegetation map units, Neldner *et al.* 1995; or species' richness, Austin & Heyligers 1989, Colwell & Coddington 1994). Another approach might be to define sampling objectives on the basis of an unbiased stratification of existing landscape attributes and spatial patchiness, using ecological definitions derived from both biotic and abiotic information sources where these are independently derived variables.

If additional sampling is impractical, then the scale at which it was intended to predict species' distribution patterns could be generalised to that at which replication of associated ecological variability is reasonably achieved with the existing set of data (e.g. like a homogeneity analysis, Bedward *et al.* 1992). Then, analyses that are limited to the existing set of compiled data need to be carefully considered in the context of the geographic areas of representativeness or bias that are inherent to the data. There are several data analytic and statistical approaches that could be devised to deal with these issues. For example, the classification of proportional sampling requirements within habitat units could be used as a basis for 'stratified random sub-sampling' from the existing set of data. The problem with this approach is that, unless the original sample is very large and very well replicated, then each subsample could result in slightly different predictive models.

If time and computing capability allowed, a randomised approach to repeated subsampling and repeated analysis could be implemented. If the sample was not so large, as is the case here, and considering the scale at which it is intended to predict plant responses, then a statistical model could be based on the existing set of samples. Depending upon the question, the sampling domain for analysis could be constrained to that region in which representative sampling were reasonable, with the results of an analysis being conditional on interpolated predictions within that domain. If a broader question is addressed, which inherently includes samples from regions that are poorly represented, then the results of such a model could be compared with just a few other models derived from representative resampling of the data. The aim here would be to test the robustness of the predictive model, given variation in sampling adequacy.

The choice of method may depend upon the preference of the analyst, or the limitations of computing repeated models and the suitability of a statistical resampling regime. Large differences in the results may indicate that predictions are biased by sampling artifacts and may have little relationship to the real landscape patterns. However, the simplest approach will be too ensure reasonable sampling adequacy for the data prior to analysis, or to generalise the questions to be addressed by the analysis to a scale which is more indicative of the sample.

2.5 Conclusions

A clear statement about the scope and precision of the ecological data intended to be used for predictive analyses of *Eucalyptus* species' occurrence in Tasmania has been possible through this analysis of sampling adequacy. The original scale at which it was intended to predict species'

distribution patterns (i.e. *ecoseries* scale of ecosystem variability), has been altered to a consideration of the ecological processes for species' distributions would approximate a classification in the *ecosection* to *ecodistrict* range (i.e. a mapping resolution of about 1:500 000 scale).

The results also support the premise that both abiotic and biotic attributes are needed to comprehensively classify landscape levels of ecological variability as a context for biophysical sampling (e.g. McKenzie *et al.* 1989, 1991), conservation reserve planning (e.g. Kirkpatrick & Brown 1994; Thackway & Cresswell 1995) and other general ecological purposes (e.g. Nix 1992; Neave & Norton 1991). This is doubly important because in many situations the ideal ecological classification of the landscape cannot be achieved due to the practical limitations of information availability (e.g. Thackway 1992; Pressey 1994; Bunce *et al.* 1996b; Host *et al.* 1996; McKenney *et al.* 1996; Smalley *et al.* 1996).

A geographic information system may introduces flexibility into survey designs and landscape classification but this does not ensure data quality. Therefore, many ecological studies would continue to benefit from the availability of contextual information published as a hierarchical-scaled atlas of environmental domains (e.g. Thackway & Cresswell 1992), incorporating commonly used combinations of landscape attributes (e.g. geology, altitude, terrain, vegetation type), in addition to process-oriented variables (e.g. climate, soil type, soil water supply, productivity potential) which are more directly relevant to biological processes.

This study effectively extended the stratified random sampling method of the GRADSECT survey design to an unbiased, retrospective analysis of ecological sampling adequacy. Other sampling designs based upon ordination, clustering or variance estimation techniques may also be applicable to an analysis of ecological representativeness (e.g. Dale 1983; Belbin 1993; Stehman & Overton 1994; Rao & Shao 1996; Faith & Walker 1996). However, the approach taken here is consistent with the concept of the landscape distribution of vegetation as an ecological continuum (Austin 1991b), and provides a simple, practical basis for variable probability allocation of samples within a systematic and objective framework.

This analysis of landscape levels of ecological variability and the corresponding sampling adequacy of an *ad hoc* set of compiled data has defined representativeness in the context of different habitats for the occurrence of eucalypt forest, irrespective of the species involved. However, where the objective is to predict a species' response from its distribution patterns (presence and absence), then the scope of the analysis would be determined by the geographical representation of individual species (Austin & Meyers 1996). A preliminary analysis of an existing set of occurrences may be needed to define or infer an environmental or bioclimatic envelope as the potential range of habitats in which a species is statistically likely to be present (e.g. Busby 1991; Carpenter *et al.* 1993; Lindenmayer *et al.* 1996). Alternatively, a comprehensive compilation of minimum levels of information could be used to define species'

geographic range limits. The potential geographic range of a species' distribution defines the broad scope for analysis and guides the sub-setting of presence and absence samples from the compiled set of ecological data. An assessment of environmental representativeness and sampling adequacy could then be undertaken within this geographic context. These problems of data design and analysis are considered in the following chapter.

3. Sampling adequacy of compiled ecological data for representing *Eucalyptus* species' geographic and environmental ranges

3.1 Introduction

The previous chapter considered the adequacy of representative sampling for a compiled set of ecological data throughout the range of eucalypt forest habitat types characterising Tasmanian sclerophyll forests. Where representative, these data were found to approximate a classification scale of eucalypt forest habitats units corresponding to the ecosystem scale of *ecosection* to *ecodistrict* (nomenclature follows Klijn & Udo de Haes 1994, previously reproduced in Table 2.1). However, the reason for compiling these ecological data was to predict *Eucalyptus* species' distributions. While the previous analysis indicated overall sampling adequacy with respect to a landscape estimate for ecological variability, some indication of sampling adequacy of individual *Eucalyptus* species' occurrences is also needed. This chapter therefore considers the sampling adequacy of the ecological dataset with respect to the distribution of individual *Eucalyptus* species in Tasmania so that (i) appropriate sampling domains can be defined for subsequent modelling, (ii) important attributes related to their geographic and environmental limits are highlighted, and ultimately, (ii) so that the success of predictions can be evaluated.

The definition of a sampling domain for individual species' distributions serves two purposes. In the first instance, it provides a basis for deciding which set of absence records to include with the set of presence records from the compiled ecological data. In the second it provides a basis for constraining species' distributions within environmental domains that accurately reflect their core and marginal occurrences. This constraint was believed necessary when modelling species' distributions because numerous zero values beyond the environmental limits of a species' occurrence tend to distort response shapes (Austin & Meyers 1996). This feature of ecological data, that creates problems in analysis and confounds interpretation of the results of predictive models, has been termed 'naughty-noughts' (Austin 1979; Austin & Meyers 1996).

In previous applications of predictive models to species' distributions, Austin and co-workers (e.g. Austin *et al.* 1994; Austin & Meyers 1996) subjectively constrained their sampling domains according to the observed response of a species along key environmental gradients, such as mean annual rainfall or mean annual temperature. The samples that fell outside the user-defined limits were not included in subsequent logistic regression models (cf. Oksanen 1997). However, this chapter seeks to define sampling domains for predictive modelling that are more objective. Species' geographic ranges, based on their parameters for location — latitude, altitude and longitude — are used as a basis for defining sampling domains. In each of these sampling

domains, the set of presence for a species, in addition to the set of absences can be sub-sampled from the compiled data and used in predictive analyses.

Therefore, the first requirement of these analyses is the definition of species' geographic and altitude ranges. One approach might be to match the environments represented by a set of samples for the presence of a species with a statistically equivalent set of environments. This is the basis of bioclimatic analyses in which the potential environmental domain for a species may be inferred from a sample of its occurrence (e.g. Nix 1986; Busby 1986, 1991; Walker & Cocks 1991; Carpenter *et al.* 1993). However, bias associated with the original set of samples may be perpetuated in defining the species' environmental domain (Neave & Norton 1991).

An alternative approach might be to compile all existing sources of information about the distribution of a species (or other taxon classification), so long as it can be positioned in geographic space. A generalised mapping scale ensures that different levels of information, from different sources, can be usefully integrated in a database and contribute to a definition of species' ranges. Data sources could comprise the validated information from museum specimen labels, in addition to ecological inventories, published results of field observations and the anecdotal records of naturalists and professional biologists (e.g. Countryman *et al.* 1981; Crovello 1981). As a result, large sets of data can be efficiently collated from diverse sources and systematically screened for errors, providing a clear indication of known distribution patterns. This initial review of existing information relating to a species' distribution could be used with an environmental domain analysis to distinguish potential habitat ranges. Alternatively, if the collation is reasonably comprehensive, it could be used directly to define a species' known geographic and environmental range, thus broadly representing its potential distribution.

A comprehensive database of species' distributions could also be applied to various ecological and phytogeographic analyses, which may be conditional on the presence of a species (e.g. Kirkpatrick 1982; Gill *et al.* 1985; Bowman *et al.* 1988; Hnautik & Maslin 1988; Myklestad & Birks 1993; Prance 1994; Rebelo 1994; Huntley *et al.* 1995; Hughes *et al.* 1996; Kelly 1996). Another application of broadly defined species' ranges is the delineation of field survey regions in which methods of stratified random sampling are implemented to ensure a representative sample for a specific purpose of ecological analysis (e.g. Austin & Heyligers 1989, 1991; Margules & Austin 1994; and see application by Neave *et al.* 1996b).

Once the known or potential geographic and altitude range for a species has been defined, then the appropriate sample of its presence and absence records can be derived from the broader compilation of ecological data. Two analyses for assessing representativeness of the species' subsample could be implemented. The first might be an assessment of ecological representativeness with respect to a classification of environments, and a subsequent assessment of sampling adequacy (e.g. Chapter 2). The second might be a more general exploratory data

analysis, considering aspects of data representativeness and environmental ranges for individual species' samples. This latter approach is the objective of this chapter.

Since the ecological response for the sample of species' occurrence in compiled ecological data is binary, then questions of sampling adequacy and environmental ranges would need to be addressed for the absence as well as the presence records within a specified sampling domain. Methods of exploratory data analysis were also used to screen each subsample of species' occurrences for errors that could bias the results of subsequent predictive modelling.

As species' distributions are influenced by biotic as well as abiotic factors of their environment, consideration was also given to the representativeness of sampling based upon a collation of published accounts of observed levels of eucalypt forest community variability in Tasmania (Duncan & Brown 1985; Kirkpatrick *et al.* 1988a). The variability inherent in the structure and composition of eucalypt forest stands provides an independent context for assessing sampling adequacy, other than approaches based on abiotic factors alone. Eucalypt forest stand characteristics, for example, reflect features of the local habitat, such as disturbance regimes and the subsequent regeneration and competition environments of a species (e.g. Florence 1996), that may not be apparent from climate and substrate factors alone.

These exploratory analyses therefore explicitly define the appropriate sampling domain for each species, and demonstrate the strengths and limitations of the existing set of presence and absence records for subsequent predictive modelling, in the context of both biotic and abiotic factors.

3.2 Methods

3.2.1 Atlas of Tasmanian *Eucalyptus* species' distributions

An atlas of the known accounts for the natural occurrence of the 29 Tasmanian *Eucalyptus* species was compiled (Williams & Potts 1996, attached with this thesis) for the purpose of delineating the appropriate sampling domains for subsequent predictive analysis of individual *Eucalyptus* species' distributions. This atlas also provides a context for assessing the geographic and environmental representativeness of the set of presence and absence records available for each species in the compiled set of ecological data (described in Chapter 2).

This atlas of *Eucalyptus* species' distributions was compiled from published and unpublished information sources for a mapping scale of 100 km^2. This reflects a classification of ecosystems at the scale of the *ecodistrict* (*sensu* Klijn & udo de Haes 1994, see Table 2.1). These data, comprising over 60 000 observations of *Eucalyptus* species' presence, were subjected to rigorous review in collaboration with B. M. Potts (screening for errors in location and taxonomic identity). The methods of collation, review and mapping were detailed in Williams & Potts (1996). The altitudinal ranges, in 100 m classes and conditional on the presence of a species, were also summarised with the distributions. These altitude data were available for analysis in this chapter

at their original resolution (i.e. ± 20 to 40 m, depending upon contour resolution given for each 1:100 000 TASMAP).

3.2.2 Assessing representativeness of *Eucalyptus* forest community variability

The floristic and structural variation of stands dominated by different *Eucalyptus* species in Tasmania were defined by Duncan & Brown (1985) for dry sclerophyll forest, and Kirkpatrick *et al.* (1988a) for wet sclerophyll forest from widespread inventory surveys. These classifications can be used to generally assess the floristic representativeness of the existing set of samples for eucalypt occurrence in the compiled set of ecological data. Although, these two classifications formally delineated forest communities at different scales, they were comparable at the scale for which floristic and structural variations were described under each classified type.

Dry sclerophyll forest vegetation was generally classified by the *Eucalyptus* species dominating the stand and by the structural formation of the stand. The structural types were described by reference to the dominant life-forms of the understorey vegetation (grassy, sedgey, heathy, shrubby), or by their situation in subalpine habitat. Informal descriptions of the variation in stand composition within each of these dry forest communities recognised distinct classes of floristic (i.e. *Eucalyptus* species' dominance and/or richness) and structural variability (e.g. low open forests and woodlands).

The wet eucalypt forest vegetation in Tasmania was formally described to the finer level of classification (Kirkpatrick *et al.* 1988a), that is descriptively defined for dry sclerophyll forest (Duncan & Brown 1985). Each wet eucalypt forest community was distinguished by the *Eucalyptus* species dominating the stand, and co-occurring tree and understory species. Additional classes of floristic variability (i.e. different combinations of *Eucalyptus* species' dominance and/or richness) could occasionally be recognised from the descriptions given with this classification. These were also taken into account for the purpose of assessing eucalypt forest community variability.

These classifications were supplemented with communities more recently recognised by Kirkpatrick *et al.* (1995). The resulting compilation of eucalypt forest community variability delineates 240 distinct eucalypt community types by dominant *Eucalyptus* species. Given the inherent continuum in forest vegetation, and to avoid the mismatching of wet or dry sclerophyll forest type within the ecological dataset, this analysis does not distinguish communities by the classification of wet or dry forest types, but rather by the structural and floristic variability associated with a particular dominant *Eucalyptus* species. The final classification therefore defines the known frequency of sclerophyll forest community variability recognised for each eucalypt species, where this species was classified as the stand dominant.

The observed level of eucalypt forest community variability in the ecological dataset was therefore estimated from a similar classification of the structural and floristic variability of

vegetation associated with each dominant *Eucalyptus* species. The sampled attributes of dominant *Eucalyptus* species, *Eucalyptus* species' richness, stand structure (height and cover) and understorey type follow the classification of biotic heterogeneity applied in Chapter 2. For consistency with the literature and problems in distinguishing these species, the community variability associated with *E. johnstonii* and *E. subcrenulata* was combined as one type (*cf.* Kirkpatrick *et al.* 1988a).

The published variability of eucalypt forest communities (Duncan & Brown 1985; Kirkpatrick *et al.* 1988a, 1995) reflects a relatively broad ecological scale, in the *ecosection* to *ecoregion* range (e.g. consider a very general approximation from an average land unit size of 10 000 ha for 240 eucalypt community types in about 2 500 000 ha of eucalypt forest as mapped by Wells (1989) and Williams (1989)). The classification of variability within the ecological dataset reflects a finer scale than the published levels, being in the *ecosection* to *ecodistrict* range (e.g. average land unit size of 2000 ha from 1335 eucalypt community types in about 2 500 000 ha of eucalypt forest, after Wells (1989) and Williams (1989)).

The eucalypt forest community variability observed from the ecological dataset was therefore standardised for comparison with the scale of variability distinguished from the published classifications. The proportion of variability ascribed to each dominant *Eucalyptus* species, relative to the total for all species, was multiplied by the 240 known forest stand types to derive an 'expected' level, for comparison with the known frequencies. A Chi-squared test, using the deviance statistic (McPherson 1990) compares the expected eucalypt forest community variability of the ecological data with the known (published) variability of forest stands. Missing sets of information relating either to the potential for dominance by a particular *Eucalyptus* species, or the potential for existence of an unclassified type of community were not used in the comparison, but were included to indicate the potential existence of these forest stand types. All classified types of eucalypt forest community variability as grouped by dominant *Eucalyptus* species were also recorded in the ecological dataset.

3.2.3 Assessing sampling adequacy of species' geographic and altitudinal ranges

The sampling domain for each species was defined as the respective known geographic and altitudinal ranges from the atlas of distributions (Williams & Potts 1996). The data subsets for each species were obtained by matching the location and altitude information from the ecological dataset to the 100 km^2 grid-cell scale and upper/lower altitude limits defined by the atlas. The rules for delineating the altitude ranges were adjusted to include the complete range for each species with additional absences at the margins to ensure clear definition of the limits to distribution for predictive modelling (Table 3.1). The relatively coarse resolution of the geographic distributions (100 km^2 cells) also contributed a number of 'trailing' absences associated with marginal occurrences. These 'trailing' absences may be viewed as the few zero values beyond the environmental limits of the species, that are expected to assist the correct

definition of species' response shapes in subsequent predictive modelling, and thereby avoid the insidious problem in analysis of 'naughty-noughts' (e.g. Austin & Meyers 1996). The resulting set of data for a species comprised all presences and a set of absences from within its geographic and altitude range, representing a classification of distributions at the *ecodistrict* scale (*sensu* Klijn & Udo de Haes 1994).

Table 3.1 Altitude ranges used in defining species' sampling domains. Central altitude tendencies (upper and lower limits) were defined within 1.5 times the interquartile range (based on data compiled for altitude distributions by Williams & Potts 1996).

Monocalyptus species	Altitude Range (m)
E. amygdalina	Sea-level to 920
E. coccifera	390 – 1310
E. delegatensis	100 – 1240
E. nitida	Sea-level to 1020
E. obliqua	Sea-level to 830
E. pauciflora	10 – 1080
E. pulchella	10 – 740
E. aff. radiata	320 – 410
E. regnans	20 – 840
E. risdonii	Sea-level to 280
E. sieberi	Sea-level to 720
E. tenuiramis	Sea-level to 700

Symphyomyrtus species	Altitude Range (m)
E. archeri	980 – 1280
E. barberi	130 – 500
E. brookeriana	20 – 720
E. cordata	20 – 680
E. dalrympleana	150 – 1150
E. globulus	Sea-level to 650
E. gunnii	500 – 1200
E. johnstonii	260 – 920
E. morrisbyi	10 – 80
E. ovata	Sea-level to 720
E. perriniana	460 – 620
E. rodwayi	60 – 1120
E. rubida	90 – 820
E. subcrenulata	550 – 1180
E. urnigera	420 – 1160
E. vernicosa	400 – 1250
E. viminalis	Sea-level to 940

The geographic representativeness of each *Eucalyptus* species' range in the ecological dataset was assessed as the number of 100 km^2 cells recording a presence or absence, relative to the known range from the atlas of distributions. Species were ranked by the proportion of cells in the ecological dataset that represented their known range. The relative proportions of presence to absence cells for each species were also compared.

Environmental representativeness of each species within the ecological dataset was generally inferred by comparison with the altitude data used to define their ranges in the atlas of distributions (Williams & Potts 1996). The Kolmogorov-Smirnov test (after Conovar 1980, and the statistical tables of Rohlf & Sokal 1969), was used to compare the continuous, relative frequency distributions of altitude for each species between the ecological dataset and the information collated for the atlas. The two sets of data were generalised at the common 100 km^2 scale by summarising identical altitudes in the same cell as a single record. Species which were represented by fewer than five samples in the empirical distribution (n) could not be compared (i.e. *E. archeri*, *E. morrisbyi*, *E. aff. radiata*, *E. vernicosa*). The atlas information was taken as the hypothetical distribution, and the representativeness of altitudes for each species in the ecological dataset (empirical distributions) was assessed by assuming these represented a random sample of the hypothetical distribution. A conservative test ($\alpha = 0.05$, for $D < D_\alpha$) was used to define the minimum vertical difference between distributions for the comparisons to be

considered the same. Rejection of the null hypothesis (that the two distributions have the same underlying function) suggests that the species in question may be inadequately represented in the ecological dataset across all or part of its known altitude range ($D \geq D_\alpha$).

The potential for the species to be adequately represented across part of its range was assessed by a Chi-squared comparison of the observed and expected frequencies of presences in 100 m altitude classes (following the method of McPherson 1990). Because these tests were correlated, a low type 1 error rate (i.e. due to the probability of incorrectly rejecting the null hypothesis under the chosen partitioning of outcomes into acceptance and rejection regions) was used ($p < 0.1$, confidence intervals for residuals: $-1.28 > r_s > +1.28$).

3.2.4 Assessing the sampling independence of altitude ranges

The sampling independence of the altitude response for the presence records of each *Eucalyptus* species, relative to the sampling range of all *Eucalyptus* species in the ecological dataset (excluding the data for the specific species being tested), was assessed using the Kolmogorov-Smirnov test. If the two distributions were the same, the species may be inadequately represented (usually indicated by relatively few samples which lead to wide confidence intervals), or altitude may not be a significant factor in its distribution. The sampling distribution for altitude across all other eucalypt occurrences in the ecological dataset was taken as the hypothetical distribution, and the distribution for individual species' presences was the empirical distribution. Using a stringent test ($\alpha = 0.01$ to 0.0001) for comparison of these distributions, rejection of the null hypothesis (that the two distributions have the same underlying function) suggests that the altitude range of the species in question is independent of the sampling range of the remaining data.

3.2.5 Assessing the complementarity of absence data

Absence information is considered to complement presence information if it comprises a different set of habitats. Together, presence and absence records more precisely define the factors which determine species' distributions. To assess complementarity, presence and absence records for each species, were treated as separate data subsets.

Test 1 — Kolmogorov-Smirnov test of altitude responses

In the first test, the Kolmogorov-Smirnov goodness-of-fit was used to distinguish whether there were significant differences in the altitude response between the two sets of occurrence information for a species (i.e. presence only or combined set of presence and absence data). Significant differences indicate that the correlation of species' distributions with environmental factors are likely to be more precise when the response is defined by both presence and absence information.

Test 2 — graphical representation of altitude responses

In the second test, the altitude response based on presence only data was graphically compared with the response derived from the combined set of presence and absence data. The probability

of occurrence in each 100 m altitude class (calculated as the proportion of presences to the number of presences and absences) was compared with the frequency of a presence (calculated as the number of presences in each 100 m altitude class relative to the total number of presences across all altitude classes) for each species. The frequency responses were conditional probabilities — conditional on the species in question being present. The relative differences between the two forms of response were generally compared for similarity in shape and magnitude. Differences in shape indicate the potential for misinterpretation of a response when analyses based on only the presence of a species are extrapolated beyond these specific conditions.

Test 3 — outlier analysis: graphical representation of climatic ranges and domains

For the third test, the environmental and geographic domains of the respective presence and absence records that define the occurrence of each species in the ecological dataset were compared in plots of annual temperature and moisture variables. The relationships between temperature and moisture variables were used because they are major factors of the environment affecting plant response at the 100 km^2 scale of comparison (e.g. Austin *et al.* 1983). These displays are also an exploratory data analysis — indicating whether further environmental restrictions may be needed to accurately represent species' occurrences for subsequent statistical modelling, or whether outliers exist that may potentially indicate data coding errors.

Test 4 — Chi-squared test of substrate fidelity

In some geographic localities, *Eucalyptus* species appear to exhibit substrate fidelity. For example, in south-eastern regions of Tasmania, *E. risdonii* is restricted to sites with skeletal soils over Permian mudstone, *E. tenuiramis* is usually associated with Permo-Triassic sediments, and *E. pulchella* is typically found on soils derived from Jurassic dolerite (Davidson *et al.* 1981). The statistical significance of substrate fidelity was assessed by comparing the frequency of species' presences on each of 14 possible substrate groupings with the combined set of presence and absence records within the respective geographic and altitude domains. These substrate groupings follow the classification of geological types used in the definition of eucalypt forest habitat units in Chapter 2. A Chi-squared test was used to assess significant differences for the null hypothesis that no substrate preferences exist ($p > 0.05$) between the presence and combined presence/absence data for each species.

In addition to distinguishing the importance of substrate in the distribution of a species and the complementarity of absence information with presence information in defining this response, the existence of substrate preferences has implications for the definition of a species' sampling domain. Potentially, the set of presence and absence information that represent a species' occurrence could be more precisely defined for the purpose of predictive analyses (Austin & Meyers 1996), by constraining geological type in addition to altitude and geographic ranges.

3.2.6 Estimating minimum sampling requirements

An assessment of minimum sampling requirements requires information about the ecological variability of habitats at the limiting *ecosection* to *ecodistrict* scale within the potential distribution of each *Eucalyptus* species (Williams & Potts 1996). However, the appropriate attributes to link information between the eucalypt atlas and published classifications of land systems (e.g. Pemberton 1986) or eucalypt forest habitat units (Wells 1989 & Williams 1989), were not available to this study. In lieu of these data, a numerical approach was developed for the estimation of ecological variability from the sample of information for the occurrence of a species, and its extrapolation into the unsampled portion of the geographic domain (Box 3.1). This simple linear, numerical extrapolation was based on the average biotic or abiotic environmental heterogeneity observed for each 100 km^2 cell, defined from the sample of occurrences (presences and absences) for each species in the ecological dataset (classifications of biotic and abiotic heterogeneity is given in Table 2.3, Chapter 2). This heterogeneity represents a classification of ecosystems at the *ecoseries* to *ecosection* scale, within the *ecodistrict* scale of the 100 km^2 cells. The minimum sampling requirement was calculated separately for the set of presences and absences for each species. Additional sampling requirements were estimated from the numerical difference between observed and expected levels.

Box 3.1. Empirical calculations developed for the extrapolation of information from the sampled domains of each eucalypt species to the unsampled domain. Estimates of sampling requirements were inferred from minimum requirements for representation and replication of extrapolated level of environmental heterogeneity.

$\hat{n}\,(\text{min}) = \hat{n}(\text{cell}) \times \overline{n(\text{env})} \times \hat{n}(\text{rep})$

Where:

$$\overline{n(\text{env})} = \frac{n(\text{env})}{n(\text{cell})}$$

Definitions:

n(env) = sampled number of different classes of environment in n(plot)

n(plot) = number of samples

n(cell) = sampled number of 100 km^2 cells

$\hat{n}\,(\text{min})$ = expected minimum number of samples

$\hat{n}\,(\text{cell})$ = potential number of 100 km^2 cells

$\overline{n(\text{env})}$ = mean number of environments

$\hat{n}\,(\text{rep})$ = estimated minimum number of replicate samples

The existing levels of environmental heterogeneity associated with the sample of presence and absence data for each species, and the number of 100 km^2 cells that comprise this sample, provide the basis for extrapolating expected levels of environmental heterogeneity into the unsampled 100 km^2 cells defined for its known distribution (Williams & Potts 1996). Multiplying the average number of environments per cell by the potential number of cells, and a minimum level of replicate samples per environment per cell (e.g. five samples), the minimum number of samples needed to represent the variability of habitats across a species' geographic range was estimated. This reflects a classification of ecosystems in the range *ecoseries* to *ecosection*. These expected levels of sampling, defined from the existing set of environments can then be compared with observed sampling levels as an indication of the representation of individual species' distributions.

This method of extrapolation assumed that the relationship between ecological variability and spatial heterogeneity was approximately linear, although the slope may differ depending on the species in question. An alternative approach to the numerical method used here would be extrapolation based on the average rarefaction curve for each species (e.g. Colwell & Coddington 1994). However, the computational requirements for this analysis were prohibitive. As a result, the linear method of extrapolation provided an indication of the trends in representation of ecological variability between species, but was likely to over-estimate sampling requirements. Therefore the ranking of additional sampling requirements for each species was used as an indication only of representativeness, standardised by relative heterogeneity defined from the sample of biophysical observations. This analysis may also have implications for indicating overall trends between species as habitat specialists or generalists, since the classification of biotic and abiotic environments provides an estimate of ecological variability relative to spatial heterogeneity.

3.3 Results

3.3.1 Assessing representativeness of *Eucalyptus* forest community variability

The comparison of eucalypt forest community variability reflects a relatively broad ecological scale, in the *ecosection* to *ecoregion* range (Table 3.2). The published accounts describe 23 *Eucalyptus* species which may occur as community dominants, and 240 different structural or floristic types. However, some species which are known to dominate vegetation associations in Tasmania appear to have been overlooked in previous surveys for the published accounts. These include stands of wet or dry sclerophyll forest dominated by *E.* aff. *radiata*, wet sclerophyll or mixed forest dominated by *E. urnigera*, dry woodlands dominated by *E. rubida* or *E. barberi*, and alpine heathlands or woodlands dominated by *E. vernicosa*. Some of these species were occasionally recorded as community dominants in the ecological dataset (Table 3.2). The rarity with which dominance by these species was recorded in the published accounts may be due to their naturally sparse occurrences (and low relative abundance) resulting in their local dominance of a site at finer scales being overlooked when assessing the mean vegetation response in 0.1 to 0.3 ha quadrats. Alternatively, it could be a reflection of the survey bias toward accessible regions, and inadequate sampling among the environments where these species may achieve some level of stand dominance.

A contrasting situation is the records of community variability associated with dominance by *E. nitida*. The published variability of stand types appears to be less than might be expected for such a wide altitude and geographic range (e.g. Williams & Potts 1996, p. 79). This low level of *E. nitida* community variability also contrasts with the greater variability of other series *Piperitae* species with narrower geographic and altitude ranges (Table 3.2). Some of these differences may be due to the relative uniformity of forest habitats in western regions (e.g. Pemberton 1989), and

others may be due to the paucity of samples from this region included in the community analyses (Kirkpatrick *et al.* 1988a).

Table 3.2 Representation of eucalypt community variability in the ecological dataset, compared with published accounts (Duncan & Brown 1985; Kirkpatrick *et al.* 1988a, 1995).

The expected level of community variability in the ecological dataset [$\hat{n}$ (comm)], was defined as the product of the observed relative variability [*n*(comm)%/100] and the 240 forest stand types [*n*(known)]. The number of observations recording the dominance of each *Eucalyptus* species in the ecological dataset is also shown [*n*(plots)]. A Chi-squared test with the deviance statistic (McPherson 1990) was used to assess differences. Missing data (.) were not used in the comparison, but indicate that such stand types may exist. The eucalypt community variability is less than expected if residuals, $r_s < -2$ (denoted by the symbol '–') or more than expected if $r_s > +2$ (denoted by the symbol, '+'), otherwise the symbol '*' is used for $0.05 > p > 0.1$, and 'ns' for $p > 0.1$. Comparisons with less eucalypt community variability than expected are considered inadequately represented for biological variability (denoted by '†').

Dominant spp.	*n*(known)	*n*(known)%	*n*(plots)	*n*(comm)	*n*(comm)%	$\hat{n}$ (comm)	r_s	sign
Stands dominated by subgenus *Monocalyptus*, series *Obliqua* species:								
E. delegatensis	38	15.8	2379	172	12.9	31	–1.1	†ns
E. obliqua	39	16.2	6360	188	14.1	34	–0 8	†ns
E. pauciflora	5	2.08	91	41	3.1	7	0.9	ns
E. regnans	5	2.08	1611	79	5.9	14	4.0	+
E. sieberi	5	2.08	120	48	3.6	9	1.8	*
Stands dominated by subgenus *Monocalyptus*, series *Piperitae* species:								
E. amygdalina	21	8.75	1951	124	9.3	22	0.2	ns
E. coccifera	14	5.83	105	51	3.8	9	–1.3	†ns
E. nitida	9	3.75	224	64	4.8	12	1.0	ns
E. pulchella	17	7.08	266	58	4.3	10	–1.7	† *
E. radiata	.	.	.	.	.	.	.	.
E. risdonii	4	1.67	15	6	0.4	1	–1.5	†ns
E. tenuiramis	14	5.83	243	53	4.0	10	–1.1	†ns
Stands dominated by subgenus *Symphyomyrtus*, series *Ovatae* species.								
E. barberi	.	.	76	26	1.9	.	.	.
E. brookeriana	4	1.67	77	24	1.8	4	0.0	ns
E. ovata	15	6.25	314	87	6.5	16	0.3	ns
E. rodwayi	9	3.75	43	22	1.6	4	–1.7	† *
Stands dominated by subgenus *Symphyomyrtus*, series *Viminales* species:								
E. archeri	3	1.25	3	3	0.2	1	–1 2	†ns
E. cordata	2	0.83	18	14	1.0	3	0.7	ns
E. dalrympleana	3	1.25	121	47	3.5	8	2 9	+
E. globulus	12	5.00	310	71	5.3	13	0.3	ns
E. gunnii	5	2.08	36	15	1.1	3	–0.9	†ns
E. johnstonii/								
E. subcrenulata	4	1.67	31	22	1.6	4	0.0	ns
E. morrisbyi	3	1.25	4	4	0.3	1	–1.2	†ns
E. perriniana	3	1.25	3	3	0.2	1	–1.2	†ns
E. rubida	.	.	7	6	0 4	.	.	.
E. urnigera	.	.	9	4	0.3	.	.	
E. vernicosa	.	.	.	.	.	.	.	.
E. viminalis	6	2.50	472	103	7.7	19	5.3	+
Totals	*240*	*100%*	*14889*	*1335*	*100%*	*240*		
						(Dev = 53; df = 22; p < 0.005)		
Eucalyptus series comparisons								
Obliquae	92	38.3	10561	528	39.6	95	0.3	ns
Piperitae	79	32.9	2804	356	26.7	64	–1.7	† *
Ovatae	28	11.7	510	159	11 9	29	0.2	ns
Viminales	41	17.1	1014	292	21.9	52	1.7	*
Totals	*240*	*100%*	*14889*	*1335*	*100%*	*240*		
						(Dev = 5.9; df = 3; p > 0.10)		
Eucalyptus subgenera comparisons								
Monocalyptus	171	71.2	13365	884	66.2	159	–0.9	†ns
Symphyomyrtus	69	28.8	1524	451	33.8	81	1.4	ns
Totals	*240*	*100%*	*14889*	*1335*	*100%*	*240*		
						(Dev = 2.8; df = 1; 0 10 > p > 0.05)		

Eucalypt community variability was summarised for taxonomic groupings of subgenera and series, in addition to species (Table 3.2). Species within the subgenus *Monocalyptus* were most frequently recorded in the literature as dominants of classified forest stands in Tasmania (171 stand types, 71%). Of these species, the series *Piperitae* species were classified as dominants in 46% of cases, with the remaining dominants being species from the series *Obliquae*. Of the classified stand types dominated by subgenus *Symphyomyrtus* species, 41% were species from the series *Ovatae*, the remaining being series *Viminales* species. Forest stands defined by the dominance of the widespread *Monocalyptus* species, *E. amygdalina*, *E. delegatensis* and *E. obliqua*, have the greatest levels of community variability (together comprising 41% of known variability). Of the *Symphyomyrtus* species, forest stands defined by the dominance of *E. ovata* or *E. globulus* also have relatively high levels of community variability (together comprising about 11% of known variability).

The standardised eucalypt community variability in the ecological dataset is consistent with the trends expected from published floristic classifications for the majority of stand types (defined by dominant *Eucalyptus* species). However, when the comparisons of variability between species are accumulated overall, for the ecological dataset, the differences are significant (df = 22, $p < 0.005$, Table 3.2). Stand types with relative community variability greater than expected (and significantly different, $p < 0.05$) are dominated by *E. regnans*, *E. dalrympleana* and *E. viminalis*, suggesting that some forest types exist that have not been formally recognised at the *ecosection* to *ecoregion* scale of classification. The relative community variability of many of the remaining stand types have been sampled in proportion to the published classifications. This comparison is especially consistent for *E. amygdalina* ($p > 0.1$), given the relative frequencies of data involved (21 stand types versus 1951 observations). At the broader taxonomic classification scale of *Eucalyptus* series and subgenera, the standardised community variability in the ecological dataset is consistent with the published summaries ($p > 0.05$).

Where fewer stand types were predicted to occur in the ecological dataset than expected from the published accounts (even where $p > 0.05$), the eucalypt community variability was not considered to be adequately represented in the dataset. Minimum levels of expected variability were more important than significance criteria when considering ecological representativeness. For example, the community variability for the dominance of six *Symphyomyrtus* species (*E. archeri*, *E. gunnii*, *E. morrisbyi*, *E. perriniana*, *E. rodwayi*) appears statistically consistent with the published accounts because stand dominance by these species is relatively uncommon. However, given the small sample sizes, even the few recognised community types are unlikely to be represented in the ecological dataset. Amongst the *Monocalyptus* species, representation of community variability for *E. risdonii* dominance also appears to be less than expected. While under-representation of community variability is also indicated for some common *Monocalyptus* species (*E. coccifera*, *E. delegatensis*, *E. obliqua*, *E. pulchella*, *E. tenuiramis*), the larger sample sizes support the findings of representativeness ($p > 0.05$).

The relative order of replication of eucalypt community variability, from unrepresented to well-represented for the classification scale of *ecosection* to *ecoregion*, is indicated in Table 3.3. The relative levels of unstandardised community variability can be high for some dominant species even when relatively few samples exist in the dataset. For example, there are four observations for *E. morrisbyi*, each representing a distinct stand type, although without replication. Conversely, there is an average of 34 observations for the types of forest stand dominated by *E. obliqua*. These replicate levels of variability appear to be higher among the *Monocalyptus* species than the *Symphyomyrtus* species.

Table 3.3 Sampling and replication adequacy for representation of eucalypt community variability in the ecological dataset. Inadequate replication was defined as fewer than 5 observations per recognised stand type [i.e. ratio of *n*(plot) to *n*(comm) from Table 3.2]. The ratio [Ratio(reps)] defines the average number of observations for each level of unstandardised variability, and therefore the level of replication in the ecological dataset. The minimum number of observations reflecting five replicates of each class of variability was extrapolated across all stand types [*n*(min)]. The additional minimum sampling [*n*(add)] required over existing levels [*n*(plot)] is indicated.

Dominating Species	*n*(plot)	*n*(comm)	Ratio(reps)	*n*(min)	*n*(add)
Unrepresented:					
E. aff. *radiata*	?	?	?	?	?
E. vernicosa	?	?	?	?	?
Inadequately Represented:					
E. archeri	3	3	1	15	12
E. perriniana	3	3	1	15	12
E. morrisbyi	4	4	1	20	16
E. rubida	7	6	1	30	23
E. cordata	18	14	1	70	52
E. johnstonii/subcrenulata	31	22	1	110	79
Poorly Represented:					
E. urnigera	9	4	2	20	11
E. gunnii	36	15	2	75	39
E. rodwayi	43	22	2	110	67
E. pauciflora	91	41	2	205	114
E. coccifera	105	51	2	255	150
E. risdonii	15	6	3	30	15
E. barberi	76	26	3	130	54
E. brookeriana	77	24	3	120	43
E. sieberi	120	48	3	240	120
E. dalrympleana	121	47	3	235	114
Reasonably Represented:					
E. nitida	224	64	4	320	96
E globulus	310	71	4	355	45
E. ovata	314	87	4	435	121
E. tenuiramis	243	53	5	265	22
E. pulchella	266	58	5	290	24
E. viminalis	472	103	5	515	43
Well Represented:					
E. delegatensis	2379	172	14	860	0
E. amygdalina	1951	124	16	620	0
E. regnans	1611	79	20	395	0
E. obliqua	6360	188	34	940	0

The results from this analysis show where sampling may be incomplete for particular species' ranges and their associated forest types (Table 3.3). For example, inadequate replication of eucalypt community variability (e.g. less than an average of 5 observations per stand type) is likely to be associated with the sets of observations dominated by *E. pauciflora*, *E. sieberi*,

E. coccifera, *E. nitida*, *E. risdonii*, *E. brookeriana*, *E. ovata*, *E. rodwayi*, and all series *Viminales* except *E. globulus* and *E. viminalis*. In general, the eucalypt community variability associated with stand types dominated by species that are less than 'reasonably represented' in the ecological dataset may not be adequately replicated for robust statistical analyses. The records for these species, however, may represent important absence information for the analysis of other species which are reasonably- or well-represented.

3.3.2 Assessing sampling adequacy of species' geographic and altitude ranges

The geographic range sizes of Tasmanian *Eucalyptus* species were defined in Williams & Potts (1996, pp.125, 149) according to their occupancy of 100 km^2 cells. This atlas represents a spatial classification of distributions at the *ecodistrict* scale (*sensu* Klijn & Udo de Haes 1994). The species' range-size distributions are strongly right-skewed (Figs. 3.1 & 3.2). This pattern is another example of the departure from the general lognormal model (Gaston 1996). The Tasmanian *Eucalyptus* species can be grouped into several different types according to their range-sizes (Williams & Potts 1996, p. 129). Each group probably represents common ecological processes such as habitat availability, habitat generalism, breadth of environmental tolerance, dispersal ability and historical factors, such as glaciation and disturbance regime (Gaston 1996). For example, the similarity in the range-size distributions of *E. obliqua* and *E. amygdalina* may reflect wide environmental tolerances which are expressed as habitat generalism (Fig. 3.1).

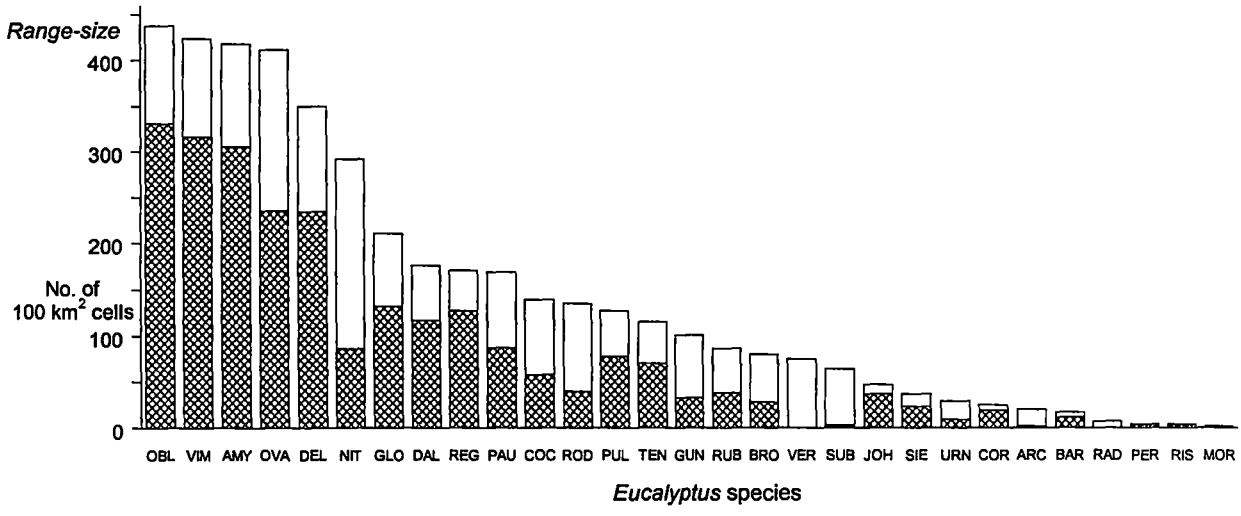
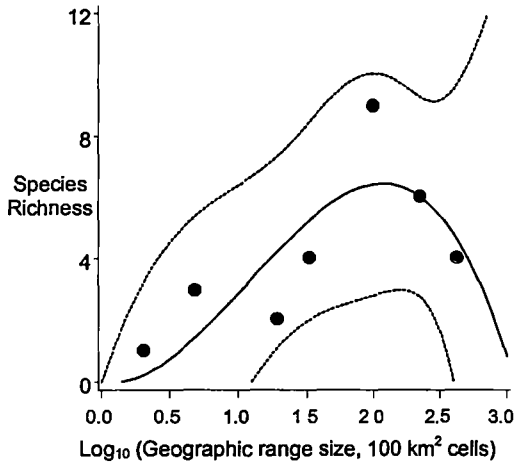


Figure 3.1 Range size distributions for the presence of Tasmanian *Eucalyptus* species in 100 km2 cells (after Williams & Potts 1996). Shading indicates the number of cells containing presence records in the ecological dataset. Codes represent the first three characters of the species' name (see Table 3.4).



Analysis of Variance
Response of species richness to $\log_{10}$(geographic range sizes)

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	146.58428	48.86143	11 906	0.0184
Error	4	16.41572	4.10393		
U Total	7	163.00000			

Root MSE	2.02582	R-square	0.8993
Dep Mean	4.14286	Adj R-sq	0.8238
C.V.	48.89900		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
MEANLOG	1	-0.766447	5.03119221	-0.152	0.8863
MEANLOG2	1	5.278467	5.54506378	0.952	0.3950
MEANLOG3	1	-1.643342	1.46134910	-1.125	0.3237

Figure 3.2 Tasmanian *Eucalyptus* species' range-size distribution for $\log_{10}$ transformed geographic range sizes (after Williams & Potts 1996). The distribution is right-skewed (see analysis of variance) and does not fit the hypothesis of a lognormal response (e.g. Gaston 1996).

Of the 29 species of *Eucalyptus* in Tasmania, and mapped in the atlas of distributions (Williams & Potts 1996), 27 are present within the ecological dataset (Fig. 3.3). Eleven of the 12 *Monocalyptus* species and 16 of the 17 *Symphyomyrtus* species were recorded as present by at least one observation. No presences for the *Monocalyptus* species *E.* aff. *radiata* or the south-western, highland *Symphyomyrtus* species, *E. vernicosa*, were recorded in the ecological dataset. This is consistent with the sampling bias for eucalypt forest habitats in lowland, eastern regions (Chapter 2). Of the 27 *Eucalyptus* species recording a presence, six are represented by less than 50 observations, contrasting with seven species which are each represented by more than 1000 observations (Fig. 3.3).

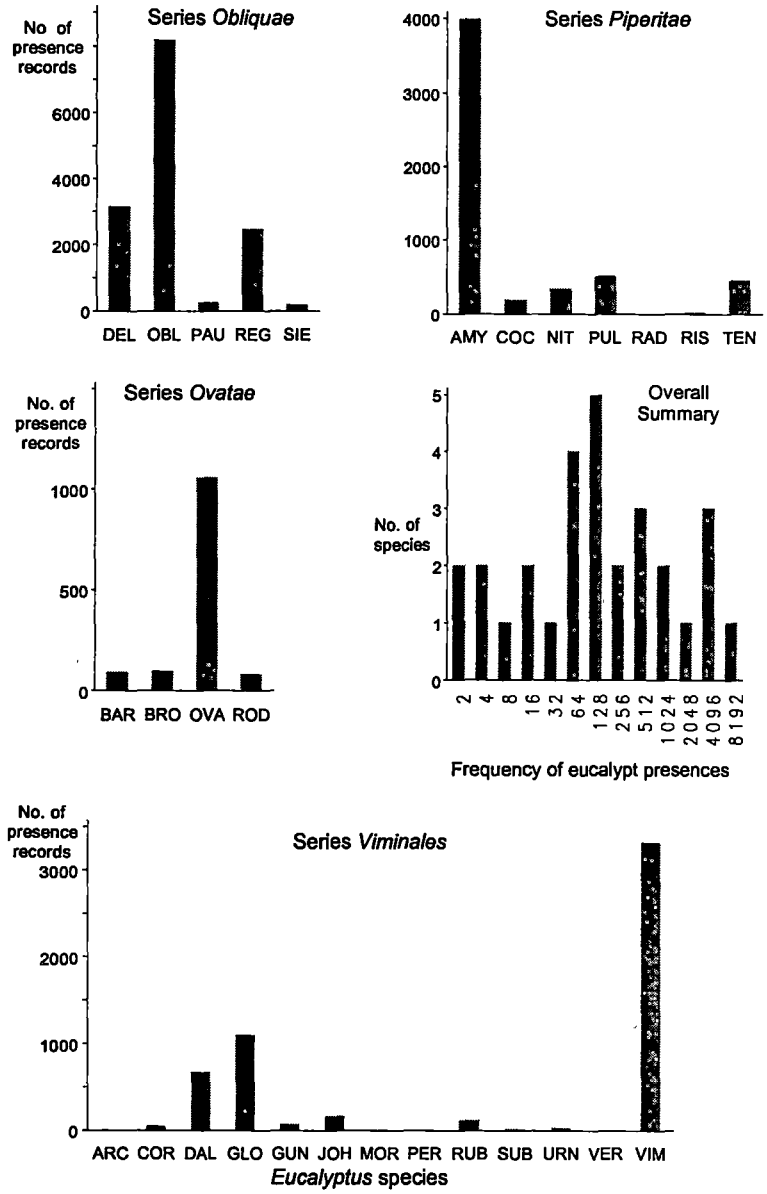


Figure 3.3. Frequency of presence for each *Eucalyptus* species by taxonomic series in the ecological dataset. Species' names are abbreviated to the first three letters of the specific name (see Table 3.5). The subgenus *Monocalyptus* comprises the series *Obliquae* and *Piperitae* and the subgenus *Symphyomyrtus* comprises the series *Ovatae* and *Viminales*.

Overall, the ecological dataset covers about two-thirds of the known geographic range for *Eucalyptus* species in Tasmania (Table 3.4). This comparison of geographic representativeness in 100 km^2 cells reflects the *ecodistrict* scale of classification. The level of representation was increased when cells recording an absence, within a species' geographic and altitude range, were also considered. The subsamples comprising both presence and absence records represent the response information available for predicting a species' distribution. Absence information exists for two species, *E.* aff. *radiata* and *E. vernicosa*, which have no presences recorded in the ecological dataset. The combined presence and absence data result in the occurrences for nine species (*E. barberi*, *E. dalrympleana*, *E. delegatensis*, *E. globulus*, *E. ovata*, *E. pauciflora*, *E. pulchella*, *E. sieberi* and *E. tenuiramis*) being completely represented across their sampling domains in 100 km^2 cells. The occurrences for a further 17 species are represented across at least

75% of their known ranges, but three species (*E. nitida*, *E. subcrenulata*, *E. vernicosa*) are very poorly represented at this classification scale, even with their set of absence records. However, a balanced sample of presence and absence information is needed to accurately represent the ecological response of species. The proportion of presence cells relative to the combined presence and absence cells for each species, indicates the potential for sampling bias in the representation of an ecological response (Table 3.4).

Table 3.4 Representation of presence (+) and absence (–), or both response types (+ & –) for each *Eucalyptus* species in the ecological dataset within its' known geographic range for the 100 km^2 grid-cell scale (from Williams & Potts 1996).

Eucalyptus species	No. cells Known from Geog-Range	No. (+) cells in Ecol-Dataset	% (+) cells in Ecol-Dataset	No. (+ & –) cells in Ecol-Dataset	% (+ & –) cells in Ecol-Dataset	Proportions: No. (+) cells to No (+ & –) cells
E. amygdalina	418	306	73.2	376	90.0	0.813
E. archeri	20	2	10.0	20	100 0	0 100
E. barberi	17	12	70.6	17	100.0	0.706
E. brookeriana	80	28	35.0	62	77.5	0.452
E. coccifera	139	58	41.7	102	73.4	0.569
E. cordata	25	19	76.0	25	100.0	0.760
E. dalrympleana	176	117	66.5	167	94.9	0.701
E delegatensis	350	235	67.1	293	83.7	0.802
E. globulus	211	132	62.6	171	81.0	0.772
E. gunnii	101	33	32.7	80	79.2	0 413
E. johnstonii	47	37	78.7	47	100.0	0.787
E. morrisbyi	2	2	100.0	2	100.0	1.000
E. nitida	292	86	29.5	128	43.8	0.672
E. obliqua	438	331	75.6	367	83.8	0.902
E ovata	412	236	57.3	331	80.3	0.713
E. pauciflora	169	87	51.5	146	86.4	0.596
E perriniana	4	3	75.0	4	100.0	0.750
E. pulchella	127	78	61.4	119	93.7	0.655
E. aff. *radiata*	7	0	0.0	7	100.0	0.000
E regnans	171	127	74.3	164	95.9	0 774
E risdonii	4	3	75.0	4	100.0	0.750
E. rodwayi	135	40	29.6	116	85 9	0.345
E. rubida	86	38	44.2	78	90.7	0.487
E. sieberi	37	23	62.2	37	100 0	0.622
E subcrenulata	64	3	4.7	31	48.4	0.097
E. tenuiramis	115	70	60.9	107	93.0	0 654
E. urnigera	29	9	31.0	28	96.6	0 321
E. vernicosa	75	0	0.0	13	17.3	0.000
E. viminalis	424	316	74.5	369	87.0	0.856
Total	*757*	*490*	*64.7*	*490*	*64.7*	*1.000*

The number of observations recording a presence or absence of a *Eucalyptus* species in the ecological dataset from within the respective sampling domains is summarised in Table 3.5. The overall geographic representation of *Eucalyptus* species is indicated by the sampling frequencies in 100 km^2 cells and the species' richness in each of these cells (Fig. 3.4).

Table 3.5 Frequency of presence (+) and absence (–) records, or both response types (+ & –), for each *Eucalyptus* species in the ecological dataset for its' known geographic range in 100 km^2 cells (after Williams & Potts 1996). The average number of records replicated per cell (calculated from cells in Table 3.4), and the proportion of presence records to absence records are also shown.

Eucalyptus species	No. (+) Records in Ecol-Dataset	Average No. (+) Records per 'Represented' cells	No. (–) Records Known in Geographic Range	Average No. (–) Records per cell in Geographic Range	Proportion (+) Records to (+ & –) Records
E. amygdalina	3991	13 0	8467	22.5	0 320
E. archeri	2	1.0	839	42.0	0.002
E. barberi	94	7.9	708	41.6	0.117
E brookeriana	95	3.4	1861	30.0	0.049
E. coccifera	185	3.2	2709	26.6	0.064
E. cordata	52	2.7	764	30.6	0 064
E. dalrympleana	674	5.8	4592	27.5	0.128
E. delegatensis	3141	13.4	7827	26.7	0.286
E. globulus	1092	8.3	4154	24.3	0.208
E. gunnii	71	2.2	2439	30.5	0.028
E. johnstonii	164	4.4	2428	51.7	0.063
E. morrisbyi	4	2.0	37	18.5	0.098
E. nitida	384	4.5	4521	35.3	0.078
E. obliqua	8183	24.7	5731	15 6	0.588
E. ovata	1055	4.5	9874	29.8	0 097
E. pauciflora	253	2.9	3189	21.8	0.074
E. perriniana	7	2.3	69	17.3	0.092
E. pulchella	462	5.9	3489	29.3	0.117
E. aff. radiata	0	0.0	270	38.6	0.000
E. regnans	2463	19.4	6838	41.7	0.265
E. risdonii	20	6.7	42	10.5	0.323
E. rodwayi	82	2 1	3976	34.3	0.020
E. rubida	116	3.1	1241	15.9	0.086
E. sieberi	182	7.9	891	24.1	0 170
E. subcrenulata	3	3.0	674	21.7	0.004
E. tenuiramis	447	6.4	3018	28.2	0.129
E. urnigera	28	3.1	1125	40.2	0.024
E. vernicosa	0	0.0	364	28 0	0.000
E. viminalis	3318	10.5	8629	23.4	0.278
Total	*14864*	*32.3*	*14864*	30.3	

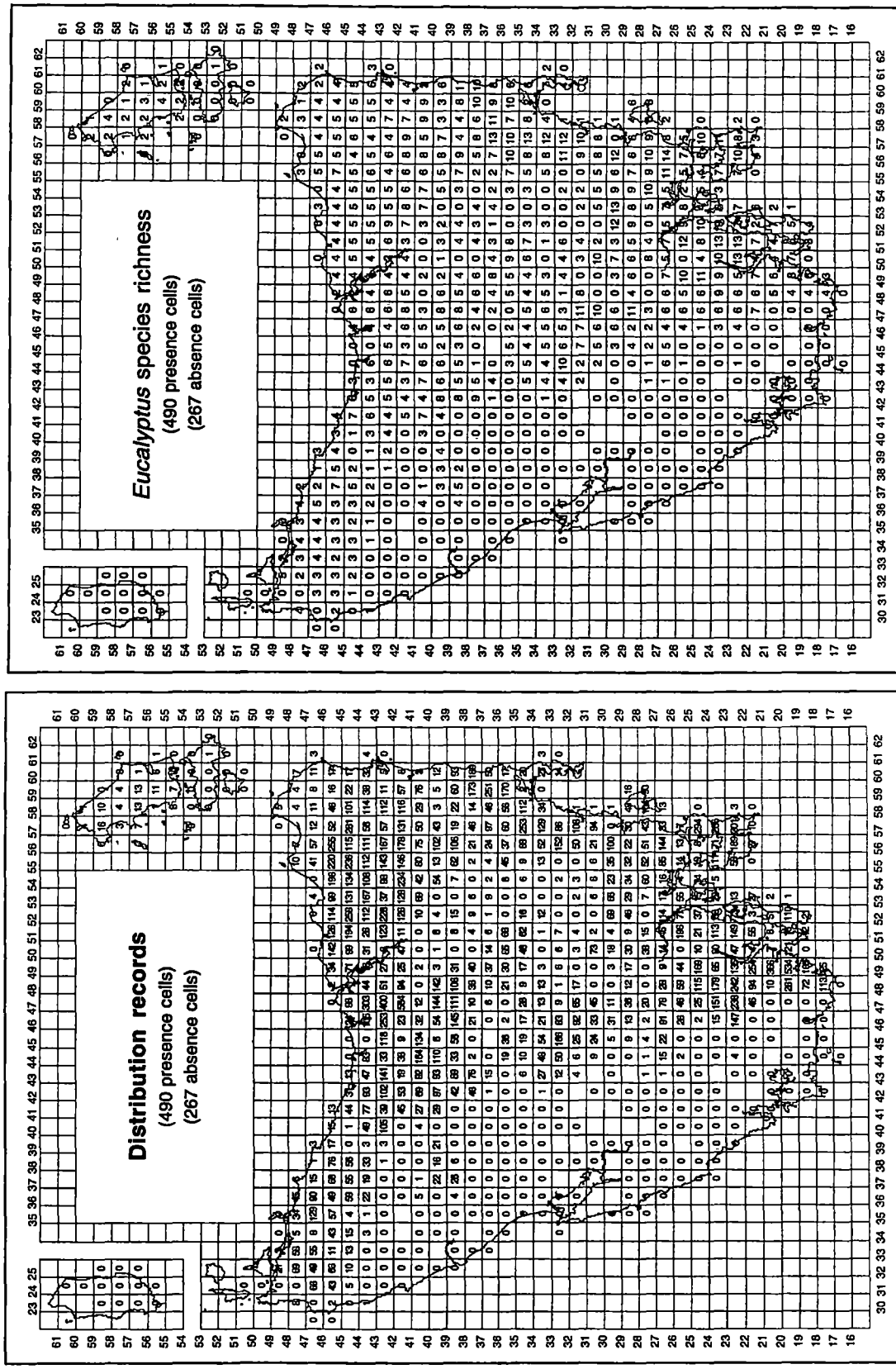


Figure 3.4 Distribution records. Location and frequency of biophysical samples (0.1 to 0.3 ha plots) for the presence of *Eucalyptus* species in the ecological dataset. Sampling frequencies are shown for 100 km^2 cells. The potential distribution of eucalypts is derived from Williams & Potts (1996). Zero values indicate locations in which eucalypts are known to occur, but for which no observations are recorded in the ecological dataset. The *Eucalyptus* species' richness associated with the set of biophysical samples in 100 km^2 cells is also shown.

The rare *Eucalyptus* species (*E. barberi*, *E. cordata*, *E. morrisbyi*, *E. perriniana*, *E. risdonii*) are frequently well-represented in the ecological dataset by their geographic ranges in 100 km^2 cells, but comprise relatively few records of a presence (Table 3.5). However, for management purposes, accurate prediction of rare species as either present or absent may have particular implications for decision making, such as whether vegetation surveys are needed prior to forestry operations (e.g. Blakesley & McDonald 1989). Therefore, the *ecodistrict* scale may be too coarse to accurately represent the ecological processes influencing the distribution of rare species. More detailed sampling of rare or uncommon species, than at levels that would otherwise represent landscape occurrences of the commoner species, may be needed. Whichever scale is the objective, species' potential geographic ranges provide a suitable context for sampling of absence information, in addition to presences.

The definition of a sampling domain avoids the problem of a large number of absence records, beyond the species' environmental range, which may reduce the precision of subsequent predictive modelling. For example, there are only 20 observations for a presence of *E. risdonii* in the ecological dataset, and a further 42 absence observations from the four 100 km^2 cells that comprises its geographic range in south eastern Tasmania (Table 3.4 & 3.5). In the case of the widespread species, *E. obliqua*, there are 8183 observations of a presence and 5731 observations of an absence within the ecological dataset, covering 367 of the 438 100 km^2 cells of its sampling domain. Delineation of sampling domains could also be applied to other criteria, depending upon the purpose of a study. For example, different taxonomic classifications (e.g. subgenera, series, sub-series) or specific ecological groupings (e.g. the geographic range of wet forest or dry forest types, mixed- or pure species stands, regional or local population ranges) could also be used to guide the selection of samples comprising presence and absence data for analysis.

The general order of representation of each species is summarised in Table 3.6. Only six species (*E. perriniana*, *E. risdonii*, *E. obliqua*, *E. cordata*, *E. johnstonii* and *E. morrisbyi*) are represented across about three-quarters or more of their known geographic ranges, and twelve others are represented across at least 50% of their ranges. Analyses of the distributions of species which are represented in smaller proportions will be very limited. However, selective analysis of species such as *E. brookeriana*, *E. coccifera*, *E. ovata* and *E. pauciflora* may be possible where these are regionally well represented. For example, *E. brookeriana* comprises two relatively disjunct distributions in eastern and north-western regions of Tasmania (see Williams & Potts 1996, p.55), but most occurrences in the ecological dataset derive from the eastern population range. Specific analyses could be restricted to the populations in this spatial domain.

Table 3.6 Relative levels of representation across each species' known geographic range (following Williams & Potts 1996). Data are summarised from Tables 3.4 & 3.5 and presented in order from unrepresented to well represented for the *ecodistrict* classification scale (100 km^2 cells).

Eucalyptus species	No. Cells Known from Geographic Range	No. (+) Cells in Ecol-Dataset	% (+) Cells in Ecol-Dataset	Average No. (+) Records per Represented Cell
Unrepresented:				
E. aff. *radiata*	7	0	0	?
E vernicosa	75	0	0	?
Poorly Unrepresented:				
E subcrenulata	64	3	4.7	3
E. archeri	20	2	10.0	1
E. nitida	292	86	29.5	5
E. rodwayi	135	40	29.6	2
E. urnigera	29	9	31.0	3
E. gunnii	101	33	32.7	2
E. brookeriana	80	28	35.0	3
E. coccifera	139	58	41.7	3
E. rubida	86	38	44.2	3
Reasonably Represented:				
E. pauciflora	169	87	51.5	3
E. ovata	412	236	57.3	5
E. tenuiramis	115	70	60.9	6
E. pulchella	127	78	61 4	6
E sieberi	37	23	62.2	8
E. globulus	211	132	62.6	8
E. dalrympleana	176	117	66.5	6
E. delegatensis	350	235	67.1	13
E. barberi	17	12	70.6	8
E. amygdalina	418	306	73.2	13
E. regnans	171	127	74.3	19
E. viminalis	424	316	74.5	11
Well Represented:				
E. perriniana	4	3	75.0	2
E. risdonii	4	3	75.0	7
E. obliqua	438	331	75.6	25
E. cordata	25	19	76.0	3
E. johnstonii	47	37	78.7	4
E. morrisbyi	2	2	100.0	2
Overall	757	490	64.7	32

The representation of altitudes in 100 m classes for all eucalypt presences in the ecological dataset was compared with the known ranges (Figure 3.5). The ecological dataset contributed about 75% of the information used to compile the known altitude ranges. The relative significance of the differences between the two sets of altitude data, although highly correlated, provides a general indication of the environmental representativeness of each species in the ecological dataset (Table 3.7). However the sampling frequencies for some species in the ecological dataset are nil, or too few for statistical comparison (i.e. $n \leq 4$, *E. archeri* and *E. morrisbyi;* $n = 0$, *E.* aff. *radiata* and *E. vernicosa*).

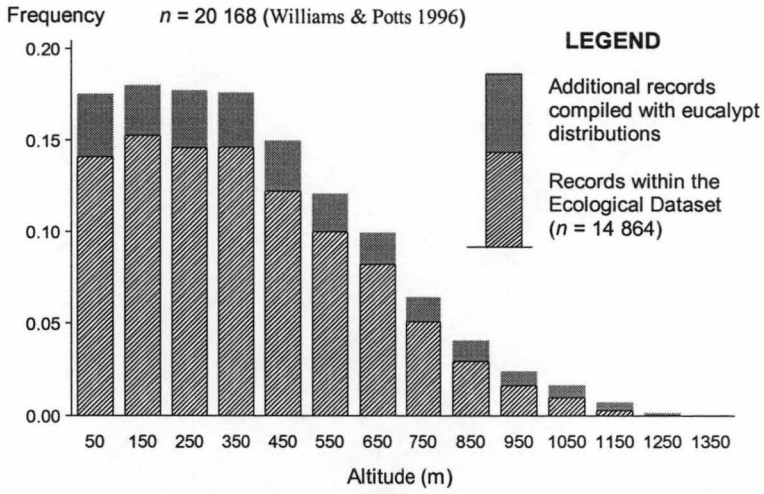


Figure 3.5 The relative contribution of altitude information from the ecological dataset to the profiles presented for the 29 *Eucalyptus* species in Williams & Potts (1996). Frequencies represent conditional probabilities for presence records of *Eucalyptus* species. Class intervals are given as mid-values.

The altitude distributions for most species in the ecological dataset are consistent with their altitude responses as collated for the eucalypt atlas (Table 3.7). The magnitude of the deviation from this hypothetical (atlas) response (D) provides an indication of the ranking of each species from well represented to poorly represented across their expected range of environments. The altitude distributions for *E. coccifera* and *E. sieberi* depart the most ($D > D_{0.05}$) from a representative trend, and *E. regnans* and *E. rodwayi* are poorly represented ($D > D_{0.1}$). The apparent representativeness for some of the remaining species ($D < D_{0.05}$) is partly due to relatively small sample sizes (n or $n_0 < 150$), resulting in wide confidence intervals. Species with a large number of samples and which are well-represented for their altitude range in the ecological dataset are *E. amygdalina*, *E. viminalis*, *E. obliqua*, and *E. delegatensis*. Species that are poorly represented, either in the frequency of observations or in the magnitude of D, are *E. brookeriana*, *E. coccifera*, *E. perriniana*, *E. risdonii*, *E. rodwayi*, *E. sieberi*, *E. subcrenulata*, and *E. urnigera.*

Table 3.7 Representation of altitude ranges for *Eucalyptus* species in the ecological dataset (empirical distribution, n) compared with their altitude responses collated for the eucalypt atlas (hypothetical distribution, n_0, see Williams & Potts 1996). The Kolmogorov-Smirnov goodness-of-fit test compares the significance of differences ($D \geq D_{0.05}$ or $D < D_{0.05}$) between the empirical and hypothetical cumulative, relative altitude distributions. Distributions which are not significantly different ($D < D_{0.05}$) are considered adequately represented (ns.). The smaller the value of the difference, D, the closer the match between distributions. The altitude in metres, χ_i, at D (the point of greatest deviation between the two distributions) is also given.

Species	n	n_0	$D_{0.05}$	D	χ_i (Altitude, m)	Representative $D < D_{0.05}$
E. amygdalina	2142	2600	0.02934	0.00991	190	ns.
E. barberi	67	79	0.16592	0.05403	280	ns.
E. brookeriana	69	125	0.16360	0.10203	150	ns.
E. coccifera	161	326	0.10703	0.15172	930	•••
E. cordata	44	67	0.20474	0.05461	150	ns.
E. dalrympleana	505	596	0.06043	0.03677	750	ns.
E. delegatensis	2171	2744	0.02915	0.01852	500	ns
E. globulus	794	1143	0.04820	0.04057	105	ns
E. gunnii	62	157	0.17248	0.09739	900	ns.
E. johnstonii	133	185	0.11776	0.03910	620	ns.
E. nitida	283	551	0.08073	0.05367	220	ns
E. obliqua	3672	4208	0.02241	0.01372	260	ns.
E ovata	691	954	0.05166	0.03107	190	ns
E. pauciflora	207	305	0.09439	0.06272	575	ns.
E. perriniana	7	11	0.51331	0.31169	540	ns
E. pulchella	337	493	0.07398	0.02250	100	ns.
E. regnans	1364	1579	0.03677	0.03477	430	ns.
E. risdonii	16	33	0.33953	0.24053	150	ns.
E rodwayi	68	186	0.16469	0.15006	660	ns.
E. rubida	98	126	0.13719	0.02041	175	ns.
E. sieberi	150	331	0.11089	0.16359	195	•••
E. subcrenulata	9	103	0.45270	0.20173	1015	ns.
E. tenuiramis	331	456	0.07465	0.03198	335	ns
E urnigera	25	59	0.27162	0.11322	830	ns.
E. viminalis	2092	2633	0.02969	0.01131	605	ns

The potential for adequate representation within a limited altitudinal range for some species ($n > 5$), is assessed in 100 m class intervals (Table 3.8). However, the 100 m intervals are too broad for some species which occur over a narrow altitude range (e.g. less than 300 m). A finer scale of classification (e.g. 50 m altitude intervals) would be needed to test the representation of *E. risdonii* and *E. perriniana*, but both were well represented for their continuous frequency distributions (Table 3.7). Species with fewer than five observations were not considered (i.e. *E. archeri*, *E. morrisbyi*, *E.* aff. *radiata*, *E. subcrenulata* and *E. vernicosa*). The Chi-squared test confirms the overall sampling inadequacy for the altitude domain of *E. coccifera* and *E. sieberi*, but demonstrates a restricted set of 100 m classes in which subsequent analyses may be valid (Table 3.8). For other species, in which the overall altitude ranges are representative of the atlas information, the chi-square test reveals some 100 m classes that are over- or under-represented. For example, the lower altitude ranges (< 100 m) of *E. brookeriana*, *E. dalrympleana* and *E. globulus* appear to be under-represented in the ecological dataset. The major class deviation is approximately consistent with the altitude for the maximum vertical difference between the hypothetical and empirical cumulative distributions (e.g. 150 m for *E. brookeriana*, and 105 m for *E. globulus*; Table 3.7). These results indicate where deviations in altitude from a

representative sample could introduce bias in analyses and where interpretation of predictions may need to account for this possibility.

Table 3.8 Representation of presences for each *Eucalyptus* species within 100 m altitude classes for the ecological dataset, compared with the atlas information (Williams & Potts 1996). The deviance statistic (*Dev*) is used with the Chi-squared test to assess differences. The symbol '***' indicates overall altitude distributions that are significantly different ($p < 0.1$, df = degrees of freedom). The sign of the standardised residual ($-1.28 > r_s > +1.28$) indicates observed frequencies in the ecological dataset that fall outside the confidence intervals for each 100 m altitude class, otherwise proportions are the same, indicated by 'ns'.

	100 m altitude class (mid-point values)														
Species	50	150	250	350	450	550	650	750	850	950	1050	1150	1250	*Dev*	*df*
E. amygdalina	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns			1.8	10
E. barberi		ns	ns	ns	ns									3.3	3
E. brookeriana	–	ns	ns	ns	ns	ns	ns	ns						9.7	7
E. coccifera				ns	ns	ns	ns	+	ns	ns	ns	–	–	***22.3	9
E. cordata	ns	ns	ns	ns	ns	ns	ns							–2.6	6
E. dalrympleana		–	ns	ns	ns	ns	ns	ns	–	ns	ns	ns		9.2	10
E. delegatensis	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	6.5	12
E. globulus	–	ns	ns	ns	ns	ns	ns		ns					6.4	7
E. gunnii						ns	ns	–	ns	ns	ns	ns		10.3	6
E. johnstonii		ns	ns	ns	ns	ns	ns	ns	ns	ns				0.4	8
E. nitida	ns	+	ns	–	ns	ns	ns	ns	ns	–	ns			10.4	10
E. obliqua	ns	ns	ns	ns	ns	ns	ns	ns	ns					0.8	8
E. ovata	ns	ns	–	ns	ns	ns	ns	ns	ns					7.5	8
E. pauciflora	ns	ns	ns	ns	ns	ns	+	ns	ns	ns	ns			6.7	10
E. perriniana					ns	ns	ns							0.0	2
E. pulchella	ns	ns	ns	ns	ns	ns	ns	ns						4.3	7
E. regnans	ns	ns	ns	ns	ns	ns	–	–	ns					13.4	8
E. risdonii	ns	ns	ns											2.1	2
E. rodwayi	+	ns	ns	+	ns	ns	ns	–	ns	ns				13.9	9
E. rubida	ns	ns	ns	ns	ns	ns	ns	ns	ns					–2.4	8
E. sieberi	–	–	ns	+	ns	+	ns							***13.3	6
E. tenuiramis	ns	ns	ns	ns	ns	ns	ns							2.7	6
E. urnigera					ns	ns	ns	ns	ns	ns	ns	ns		3.3	7
E. viminalis	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns				4.2	9

3.3.3 Assessing the sampling independence of altitude ranges

The relative independence of the altitude distributions for individual *Eucalyptus* species (n_t), within the ecological dataset, was assessed by comparison with the sampling distribution for all other *Eucalyptus* species (Table 3.9). Even with a relatively small number of samples in some cases ($7 < n_t < 100$), the observed altitude frequency distribution of each *Eucalyptus* species was significantly different (for at least $D_{0.01}$) from that which may be expected given the overall sampling distribution. These differences demonstrate an ecological preference for a particular set of environments, independent of the sampling distribution. However, some species that are represented by relatively few presences (leading to wide confidence intervals) compared with the sampling distributions (e.g. *E. perriniana* and *E. subcrenulata*), are not sufficiently sampled to accurately reflect their altitude response in this comparison.

Table 3.9 Sampling independence of the altitude range for *Eucalyptus* species.
The Kolmogorov-Smirnov goodness-of-fit test compares the significance of differences ($D \geq D_\alpha$ or $D \leq D_\alpha$) between the cumulative, relative altitude distributions for individual *Eucalyptus* species from the ecological dataset (empirical distribution, n_i; for $n_i > 5$) and the sampling range for all other eucalypts within the ecological dataset (hypothetical distribution, $n_0 = 14\,864 - n_i$). The estimated frequency distribution is generally compared with the observed frequency distribution for $\alpha = 0.0001$, a more restrictive test in the case of large samples ($n > 100$). The significance levels at which the distributions are different are indicated by '***' for the 0.01% level ($D_{0.0001}$), or '**' for the 0.1% level ($D_{0.001}$), or '*' for the 1.0% level ($D_{0.01}$). The altitude in metres, χ_i, at D (the point of greatest vertical deviation between distributions) and overall summary statistics for altitude (median, mean and standard deviation) are given to the nearest 5 m.

Species	D_α	D	n_i	χ_i (altitude, m)	Median	Mean	Std Dev
E. amygdalina	0.03522	0.11499***	3991	200	210	265	195
E. barberi	0.22952	0.25457***	94	480	260	275	90
E. brookeriana	0.22711	0.37742***	96	280	475	450	160
E. coccifera	0.16360	0.77459***	185	590	860	875	185
E. cordata	0.27034	0.30141**	52	330	480	425	185
E. dalrympleana	0.08571	0.56225***	674	430	620	625	175
E. delegatensis	0.03970	0.55406***	3141	450	660	655	195
E. globulus	0.06734	0.24233***	1092	400	210	220	145
E. gunnii	0.26409	0.77891***	71	540	940	885	215
E. johnstonii	0.17376	0.60152***	164	530	700	655	175
E. nitida	0.11356	0.22381***	384	140	160	240	225
E. obliqua	0.02460	0.16593***	8183	400	210	245	160
E. ovata	0.06851	0.22710***	1055	230	160	205	175
E. pauciflora	0.13990	0.40180***	253	430	580	545	245
E. perriniana	0.57581	0.71515*	7	450	520	520	30
E. pulchella	0.10353	0.13535***	462	400	260	285	155
E. regnans	0.04484	0.17814***	2463	180	320	340	155
E. risdonii	0.49758	0.67006***	20	150	100	110	35
E. rodwayi	0.24574	0.40425***	82	355	515	505	200
E. rubida	0.20661	0.31675***	116	230	460	440	170
E. sieberi	0.12065	0.14388*	182	590	360	330	175
E. subcrenulata	0.74175	0.86995***	9	650	920	910	160
E. tenuiramis	0.10502	0.17918***	449	560	240	260	155
E. urnigera	0.42053	0.79313***	28	595	845	805	175
E. viminalis	0.03863	0.04988***	3318	440	270	310	215
All Eucalypts (n_0)	0.01825	na	14864	na	280	340	245

Two species, *E. sieberi* and *E. cordata*, have frequency distributions for altitude which differ from sampling distribution at the lower levels of significance (Table 3.9). In the case of *E. sieberi*, the level of significance is lower than for other species with similar sample sizes, indicating that its altitude occurrences have some coincidence with the overall sampling distribution. However, previous analyses (Tables 3.7 & 3.8) indicated that the representation of *E. sieberi* in the ecological dataset is inadequate compared with its known altitude range. As a result, the relative sampling independence, or altitude preference, of *E. sieberi* cannot be accurately concluded from this analysis. The situation is more complex in the case of *E. cordata*, because its altitude response is confounded by the existence of two distinct morphs — a small round-stemmed under-storey shrub typical of drier habitats in lowland, eastern regions, and a taller square-stemmed canopy tree found in moister habitats of mid-altitude ranges in western regions (Potts 1989). The separate responses of these morphs needs to be considered, before concluding a lack of preference for altitude.

The relative deviations (D) of some *Eucalyptus* species from the overall sampling distribution for altitude (especially where sample sizes are reasonable large, $n_i > 150$), indicates the relative strength or weakness of their preferences for a particular habitat type. For example, the species

with the largest deviations from the sampling distribution (e.g. $D > 0.5$) occur in the highland or subalpine habitats (*E. coccifera*, *E. dalrympleana*, *E. delegatensis* and *E. johnstonii*). Their deviation from the sampling range for altitude is exacerbated by the predominance of low altitude records in the ecological dataset (see Fig. 3.6). Whilst all lowland species are significantly different from the sampling distribution for altitude, the smallest deviation occurs for *E. viminalis* ($D < 0.05$). Other typically lowland species also have relatively small deviations from the sampling distributions for altitude (e.g. $D < 0.25$ for *E. amygdalina*, *E. globulus*, *E. nitida*, *E. obliqua*, *E. ovata*, *E. sieberi* and *E. tenuiramis*). These differences indicate that most species have specific habitat preferences that are to some extent correlated with altitude, even if these ranges are not completely representative of their occurrences in the landscape.

3.3.4 Assessing the complementarity of absence data

The relative complementarity of the absence information from the sampling domain for each species in the ecological dataset is demonstrated by comparing the altitude distributions for the presence data with the absence information (Table 3.10). Significant differences between the two distributions indicate that the absence information contributes additional information about the occurrence of a species. The results indicate that most species occupy a specific altitude habitat (at the 0.01% level) within their environmental domains (geographic and altitude ranges from Williams & Potts 1996). The exceptions are *E. barberi*, *E. perriniana*, *E. pulchella*, *E. risdonii*, *E. rodwayi* and *E. rubida*. *Eucalyptus pulchella* and *E. rodwayi* differ at the 0.1% level, but for *E. barberi* the response of the presences are significantly different from the absences only at the 1% level. For some of these species, the similarity of altitude distribution between presence and absence records may be explained by inadequate sampling. However, for other species such as *E. pulchella*, altitude and therefore temperature may not be as significant as other factors (e.g. water balance and substrate) in determining distribution patterns within its geographic and altitude domain.

Table 3.10 Relative contribution of absence information to the altitude response of *Eucalyptus* species represented in the ecological dataset. The set of presence and absence data for each species were defined by their geographic and altitude domains (see Williams & Potts 1996). The Kolmogorov-Smirnov goodness-of-fit test compares the significance of differences ($D \geq D_\alpha$ or $D \leq D_\alpha$) between the cumulative, relative altitude distributions for presence records (empirical distribution, n_i; for $n_i > 5$) and absence records (hypothetical distribution, n_0). In the case of large samples ($n_i > 100$), the critical distance, D_α ($\alpha = 0.0001$), is a more restrictive test. The significance levels at which the distributions are different are indicated by '***' for the 0.01% level ($D_{0\,0001}$), or '**' for the 0.1% level ($D_{0\,001}$), or '*' for the 1.0% level ($D_{0\,01}$). The altitude in metres, χ_i, at D (the point of greatest vertical deviation between distributions) is also given.

Species	n_i	n_0	$D_{0\,0001}$	D	χ_i (altitude, m)
E. amygdalina	3986	8369	0.03525	0.19420***	190
E. barberi	93	519	0.22952	0.19950**	360
E. brookeriana	95	1745	0.22831	0.32813***	280
E. coccifera	181	1931	0.16360	0.43681***	750
E. cordata	52	735	0 30859	0.39017***	330
E. dalrympleana	671	4558	0.08571	0.29066***	450
E. delegatensis	3145	7601	0 03970	0.63648***	470
E. globulus	1093	3986	0.06737	0.15877***	400
E. gunnii	69	2023	0.26409	0.44833***	870
E. johnstonii	155	1196	0.17647	0.59445***	530
E. nitida	335	4585	0.11356	0.16717***	50
E. obliqua	8197	5450	0.02460	0.32825***	390
E. ovata	1052	9401	0.06864	0.20773***	90
E. pauciflora	253	3161	0.13990	0.19833***	430
E. perriniana	7	32	0.84107	0.25000(ns)	550
E. pulchella	463	3384	0.10353	0.09423**	135
E. regnans	2462	6727	0.04485	0.21870***	180
E. risdonii	20	40	0.49758	0.30000(ns)	150
E. rodwayi	82	3955	0.24574	0.22992**	340
E. rubida	115	1172	0.20751	0.14451(ns)	560
E. sieberi	182	859	0.16495	0.17191***	550
E. tenuiramis	448	2951	0.10525	0.11046***	550
E. urnigera	28	450	0.42053	0.50698***	820
E. viminalis	3319	8509	0.03863	0.08943***	340

The differences in ecological interpretation that may result from analyses based on presence-only data compared with analyses based on presence/absence data, can be graphically demonstrated for species' responses to altitude in 100 m classes. For example, the lowland species *E. obliqua* has a truncated response toward low altitudes that is only slightly accentuated for the probabilities of occurrence when compared with the conditional frequencies (Fig. 3.6a). However, in the case of *E. regnans*, a slightly skewed response toward low altitudes is indicated by the conditional frequencies, but the optimum of this response shifts toward higher altitudes when viewed as the probabilities of occurrence (Fig. 3.6b).

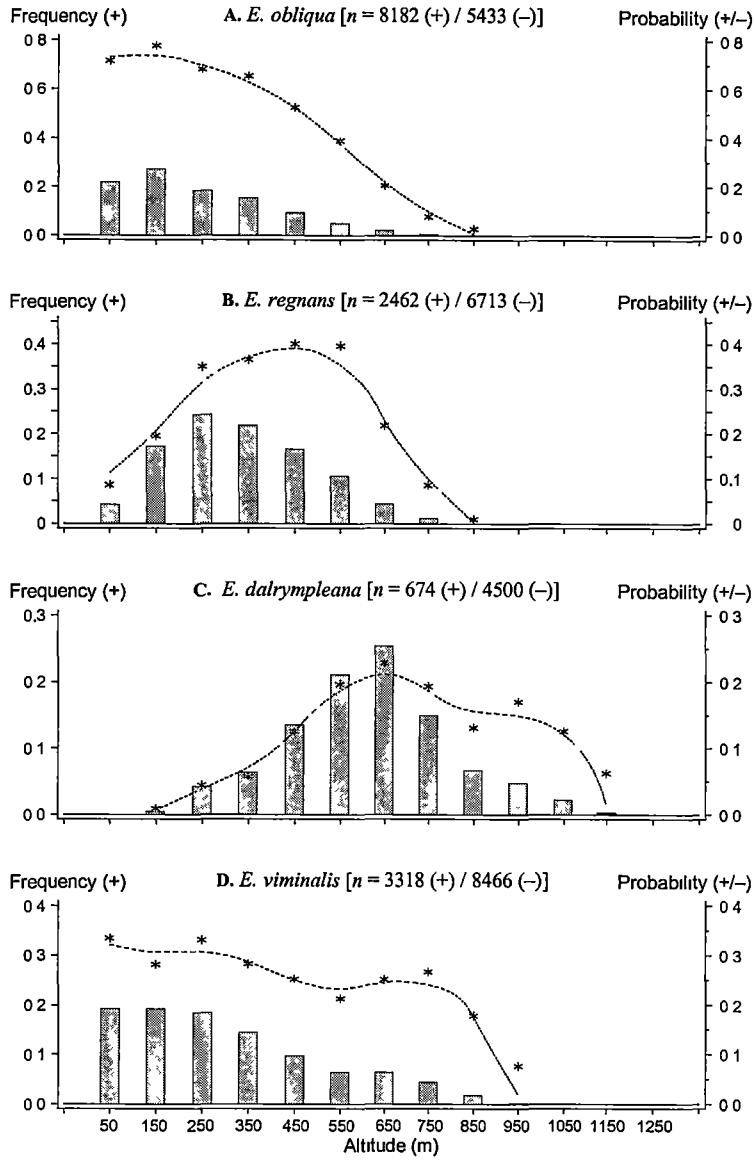


Figure 3.6 Comparison of analyses based on presence-only data or presence/absence data.

Conditional frequencies for the presence of a species in 100 m altitude classes are given in histograms (+).

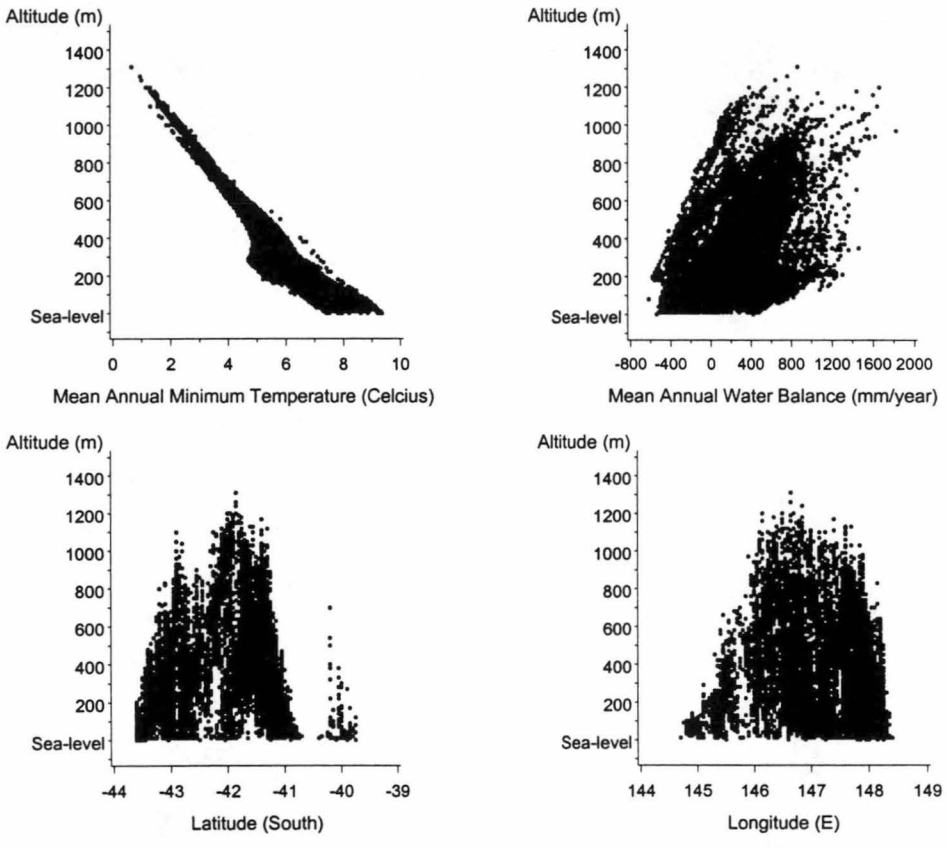
Probability of occurrence [presence relative to presence & absence (+/–)] are indicated by '*', and a running mean displays the general form of response (summed weights of 0.5 to the observation and 0.25 to each adjacent class, or for defining the end value of response — summed weights of 0.75 to the observation and 0.25 to the adjacent class).

Occurrence data for each species were sub-sampled from the ecological dataset within their respective geographic and altitude ranges (after Williams & Potts 1996).

The probability of occurrence response also highlights potential problems of taxonomic identification between species, that are not so well clarified by the presence-only response. For example, the clinal species *E. dalrympleana* and *E. viminalis* are frequently difficult to distinguish at intermediate altitude ranges in Tasmania (Phillips & Reid 1980). These difficulties are apparent in the generalised response to altitude at around 650 m (Fig. 3.6c, d). The probability of occurrence indicates probable misidentification of *E. dalrympleana* which may have been recorded as high-altitude occurrences of *E. viminalis*. These responses could be considered together in subsequent analyses, consistent with their natural cline.

Since regional climates can be reasonably well approximated to location (e.g. Hutchinson 1989, Sturman & Tapper 1996), the geographic and altitude ranges which define the sampling domains for each *Eucalyptus* species, can be redefined as climatic domains at the 100 km^2 scale. Correlations between location (latitude, longitude and altitude) and climate (water and temperature variables) are demonstrated for the range of observations in the ecological dataset (Fig. 3.7). The three indirect gradients of latitude, longitude and altitude are poorly correlated

with each other, but are strongly correlated with the more direct gradients of climate, supporting their use in defining sampling domains. Temperature and altitude are particularly well correlated, especially with increasing elevation (Fig. 3.7a). The relationships between location and climate also reflect the effect of continentality and inland mountain ranges, with wide-ranging values in mid-latitude and mid-longitude regions (Fig. 3.7a, b). The relationship between temperature and longitude reveals how altitude, to some extent, mimics latitude, and how cooler habitats at higher altitudes approximate a southerly shift in location. There is a reasonable correlation between longitude and net water balance, consistent with the prevailing westerly direction of rainfall events in Tasmania. Annual net water deficits are unusual in western regions (e.g. Longitude < 146 East), but are increasingly typical of eastern regions. Therefore, species' geographic and altitude ranges, which can be simply compiled and mapped (e.g. Williams & Potts 1996), provide an effective surrogate for their environmental domains.



PEARSON CORRELATION MATRIX:

	Mean annual minimum temperature	Mean annual net water balance
Altitude	-0.94149	0.43968
Temperature		-0.45107

	Latitude	Longitude
Altitude	0.02932	0.05624
Latitude		-0.13863

Figure 3.7a The relative contributions of altitude (metres above sea-level), latitude (° South) and longitude (° East) to a description of the climatic attributes for temperature and moisture. Temperature is given as mean annual minimum (°C), moisture is given as mean annual net water balance (precipitation minus evaporation, mm/year), for the observations in the ecological dataset (n = 15 611).

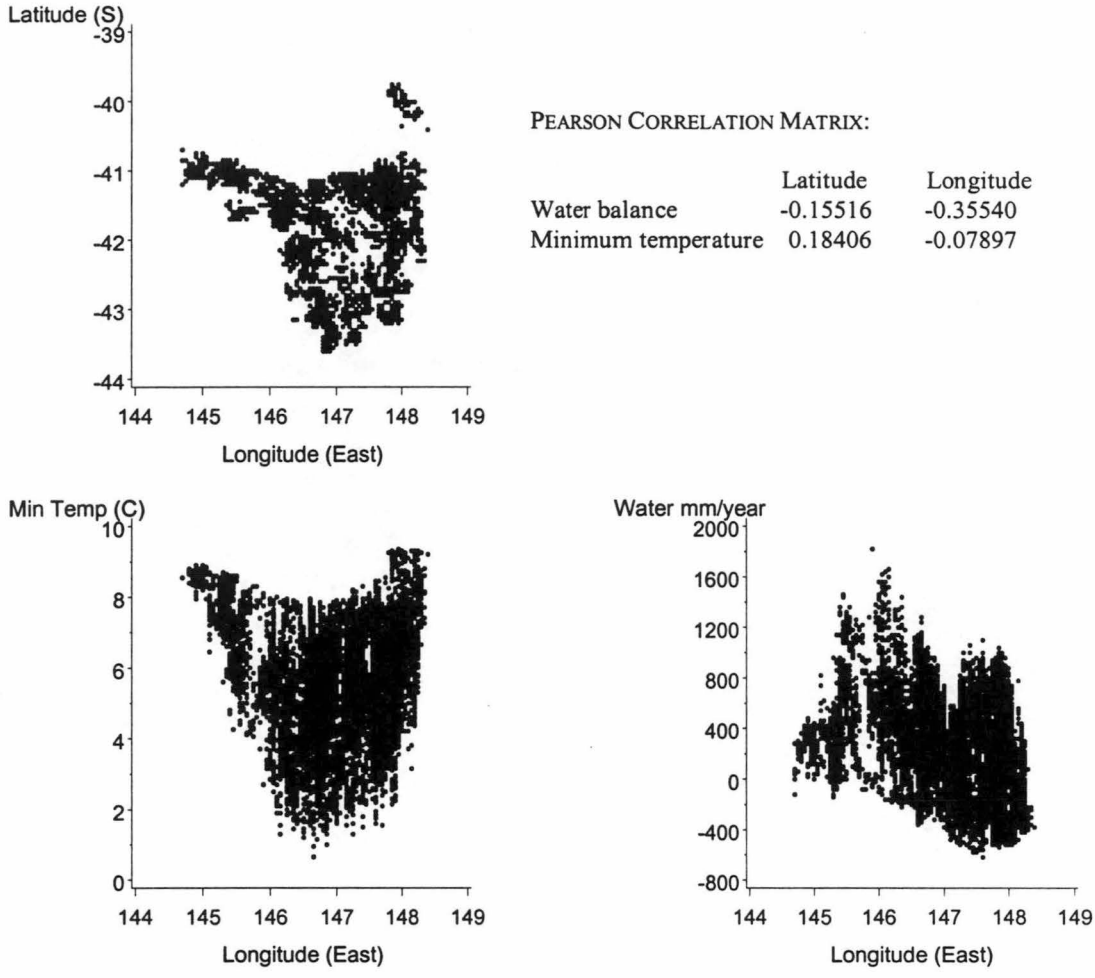


Figure 3.7b The relative contributions of latitude (° South) and longitude (° East) to a description of the climatic attributes, mean annual minimum temperature (°C) and mean annual net water balance (precipitation minus evaporation, mm/year), for the observations in the ecological dataset (n = 15 611).

These climatic data enable the environmental envelopes that represent the geographic and altitude ranges of each species to be more clearly defined in term of gradients that directly influence plant physiological responses (after Austin & Smith 1989). A more concise indication of the environmental representativeness of species' presence and absence records in the ecological dataset is thus possible from the graphical display of their potential climatic domains. For example, the potential for ecological differences among the geographically-disjunct occurrences of *E. brookeriana* are clarified by the environmental disjunction of these habitats in the climatic envelope for water and temperature (Fig. 3.8A). The bias toward representation of eastern populations of *E. brookeriana* in the ecological dataset is also graphically demonstrated. Similar climatic profiles for *E. globulus* indicate that this species is well represented for the set of absence records and reasonably represented for the set of presence data (Fig. 3.8B). These climatic envelopes demonstrate the utility of geographic and altitude ranges for defining the sampling domains of individual species' distributions, and therefore constraints on the absence data to the relevant subsets for analysis.

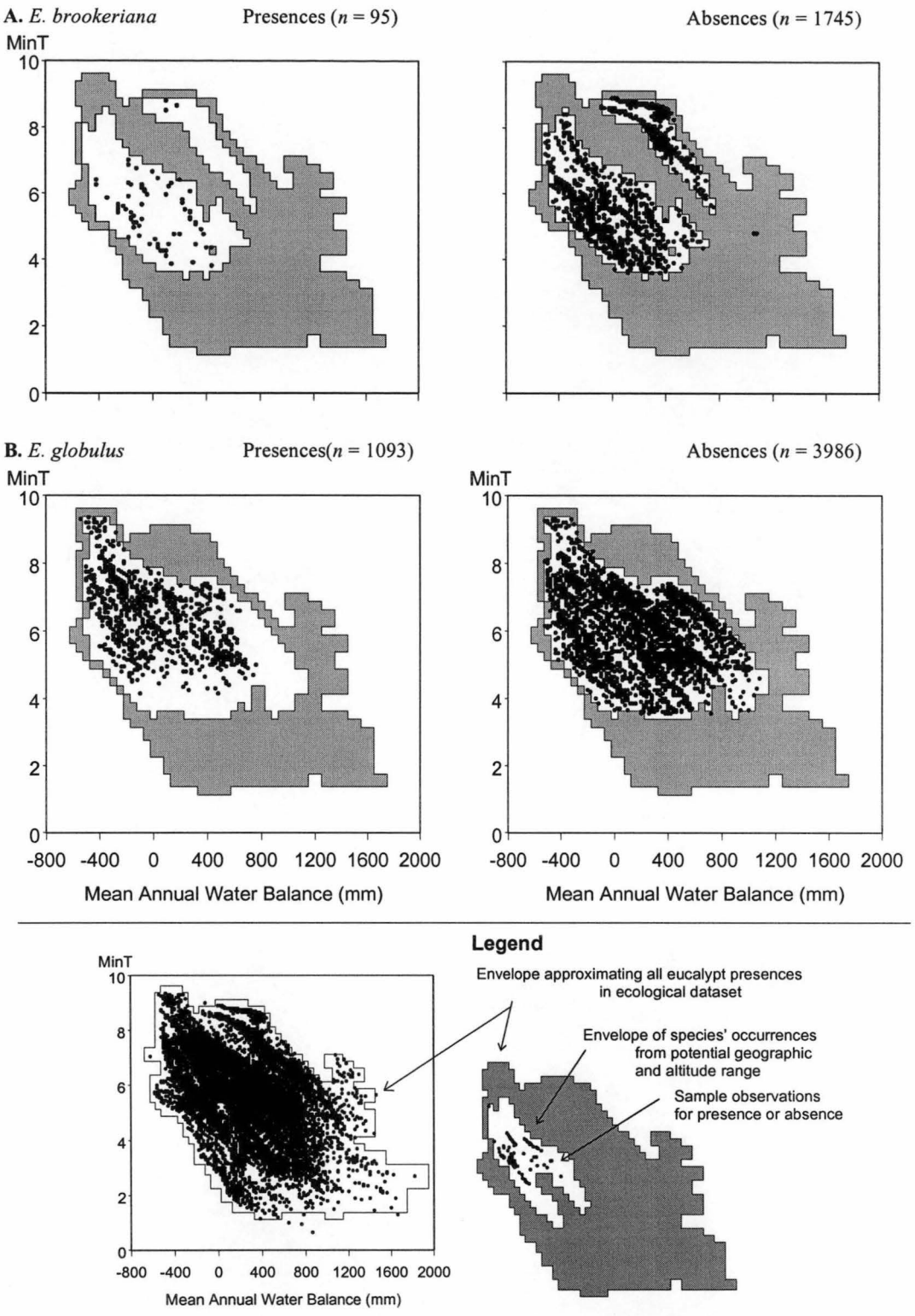


Figure 3.8 Example of climatic envelope for *Eucalyptus* species' occurrences (presences and absences) within the ecological dataset for mean annual minimum temperature (MinT, °C) and mean annual water balance (precipitation minus evaporation, mm/year). The climatic envelope of the sampling domain is defined from the geographic altitude ranges (after Williams & Potts 1996).

The display of data within univariate plots is also an important exploratory tool for highlighting environmental outliers in a species' distribution (e.g. Fig. 3.9). These outliers may be indicative of errors, such as incorrect taxonomic identification or incorrectly located samples, leading to spurious correlations

between species' occurrence and environment which may unduly influence subsequent analyses. For example, a far-outlying record for *E. coccifera* on the water gradient (Fig. 3.9A) may be a highland occurrence of *E. nitida*, since it occurs in the east-west intergrade zone between these two species (Shaw *et al.* 1984). However, incomplete representation of either *E. coccifera* or *E. nitida* across their geographic ranges (Tables 3.4 & 3.5) and the skewed relationships with the water gradient indicate that these outliers may be important occurrences for defining the potential limits of these species' distributions. Observations which are outliers on two or more independent gradients (e.g. water and temperature, Fig. 3.8) are likely to have much more influence on the results of a statistical model, checks for which are an important aspect of analysis (e.g. Cook & Weisberg 1982, McCullagh & Nelder 1989).

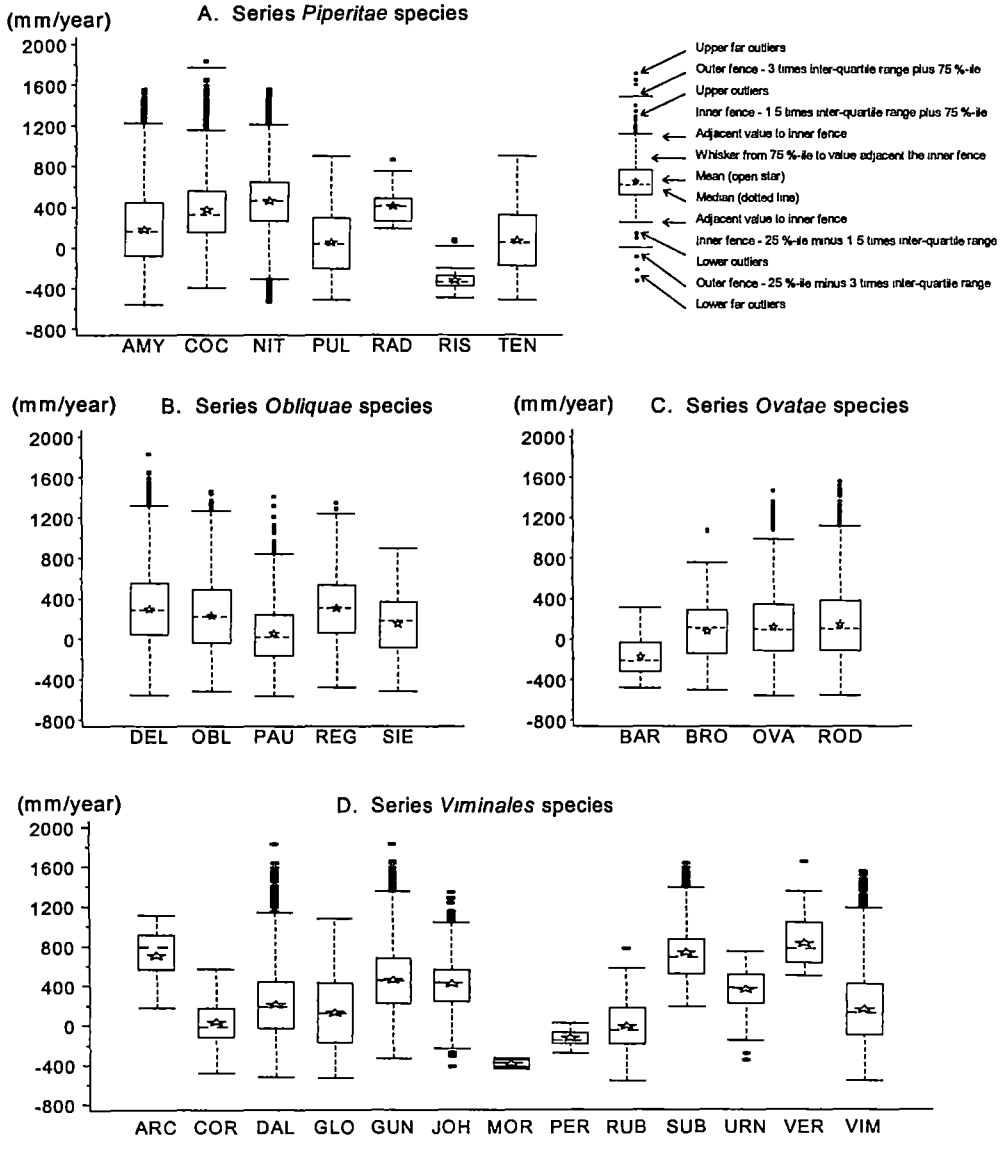


Figure 3.9 Box-plots for the univariate distribution of the net water balance variable (precipitation minus evaporation, mm/year) for *Eucalyptus* species' occurrences (presences & absences) within the ecological dataset. Occurrences defined from the geographic and altitude domains (after Williams & Potts 1996). Species are grouped by taxonomic series, labels refer to the first three letters of the species' name (see Table 3.4).

The potential for parent rock type as an additional environmental constraint for defining a species' sampling domain was also investigated (Table 3.11). Since substrate types (e.g. texture

class or nutrient index) can be inferred from parent rock type (e.g. Nix *et al.* 1992), fidelity could be related to the combined effects of moisture regime, nutrient status, or physical factors affecting plant root exploration of the soil. For example, *E. barberi* is consistently found on substrates underlain by Jurassic dolerite (e.g. McEntee *et al.* 1994), and other species such as *E. tenuiramis* may occur in association with Permo-Triassic sediments, in addition to Jurassic dolerite (e.g. Davidson *et al.* 1981; Duncan & Brown 1985). Therefore, the patterns of species' occurrence (presence and absence) by parent rock types might indicate substrate fidelity, and the potential suitability of further constraints to the sampling domain for subsequent analysis. However, some species may not be exhibiting preferences for particular substrates, but rather occurring on those substrates which are available within their climatic domains.

Table 3.11 Relative substrate fidelity among *Eucalyptus* species in the ecological dataset.
Chi-square test of differences between frequency distribution of presence records relative to presence and absence records within the respective geographic and altitude sampling domains. Species which show no significant fidelity for substrate ($p > 0.05$) are indicated by the superscript 'ns'. Comparisons are only made for substrate classes which exist in a sampling domain. Significant differences within species groups are indicated by the sign of the residuals (where $-2 > R_s > +2$), otherwise the responses are the same, 'ns' for $p > 0.05$. Substrate categories (grouped from 1:500 000 Geological Map of Tasmania, Department of Mines 1976): Cambrian sediments (CS), Cambrian volcanics (CV), Devonian granite (DG), Jurassic dolerite (JD), Mathinna beds (MT), Ordovician limestone (OL), Ordovician quartzites (OQ), Permo-Triassic sediments (PT), Precambrian metamorphics (PM), Precambrian sediments (PS), Quaternary deposits (QD), Quaternary till & talus (QT), Tertiary basalt (TB), Tertiary sediments (TS).

Geological type	CS	CV	DG	JD	MT	OL	OQ	PT	PM	PS	QD	QT	TB	TS
Total Sample (*n*)	200	277	1631	4697	1493	46	278	1558	157	436	427	55	167	406
Subgenus *Monocalyptus* species, series *Piperitae*:														
E. amygdalina	ns	ns	ns	-	+	-	+	-	+	ns	+	-	-	+
n	36	92	467	1233	632	4	157	479	89	165	236	11	20	365
E. coccifera	ns	ns		+		ns	ns	-	ns	ns	ns	ns	-	
n	0	1	.	155	.	0	6	8	0	0	1	10	0	.
E. nitida	-	ns	+	-	+	ns	ns	-	-	+	+	ns	+	
n	3	3	34	54	14	3	10	21	0	107	61	3	22	.
E. pulchella			ns	+				-			-		ns	ns
n	.	.	0	414		.	.	43			3	.	1	1
E. risdonii (ns)				ns				ns						ns
n	.	.	.	0	.	.	.	20	.	.	.	.	.	0
E. tenuiramis			+	-				+			+		ns	ns
n	.	.	11	234		.	.	179	.	.	21	.	2	1
Subgenus *Monocalyptus* species, series *Obliquae*:														
E. delegatensis	-	-	ns	+	-	-	-	-	+	-	-	+	+	-
n	22	22	371	1917	169	14	51	181	72	27	13	160	125	1
E. obliqua	+	+	-	-	ns	+	ns	+	ns	+	-	-	-	-
n	364	251	806	2676	909	91	188	1471	75	951	165	3	91	156
E. pauciflora	ns		ns	ns	-		ns	ns	ns		+	+	+	-
n	2	.	13	128	7	.	1	33	2		24	24	16	3
E. regnans	ns	-	+	-	+	ns	-	ns	-	-	-	ns	-	-
n	50	35	571	825	448	18	21	338	10	105	19	4	13	5
E. sieberi			-	-	+			ns			ns			ns
n	.	.	38	18	114	.	.	7	.	.	4	.		1
Subgenus *Symphyomyrtus* species, series *Ovatae*:														
E. barberi (ns)				ns				ns			ns			
n	.	.	.	93	.	.		0	.		0		.	.
E. brookeriana	-		ns	+	-			ns		-	-		ns	
n	2	.	2	83	1		.	6	.	0	0	.	1	.
E. ovata	+	ns	-	ns	-	ns	-	ns	ns	ns	+		+	+
n	53	6	35	404	57	1	1	151	3	66	179	.	27	69
E. rodwayi			-	ns	-			+	ns		+	ns	ns	ns
n	.	.	1	51	1	.		17	0		10	0	1	1
Subgenus *Symphyomyrtus* species, series *Viminales*:														
E. archeri				ns										
n	.	.	.	2	.	.	.		.	.	.			
E. cordata (ns)				ns				ns			ns			
n	.	.	.	44	.	.	.	8	.	.	0		.	.
E. dalrympleana	ns	-	-	+	-	-	-	ns	ns		ns	+	ns	ns
n	6	2	14	484	16	0	2	56	6	.	5	55	25	0
E. globulus			+	ns	-	ns		ns	ns	-	ns		ns	ns
n	.	.	53	750	10	0	.	248	0	0	28	.	1	3
E. gunnii		ns	-	ns	ns		ns	+	ns			ns	ns	
n	.	0	0	41	0	.	4	12	5		.	5	2	.
E. johnstonii (ns)	ns			ns				ns		ns	ns			
n	1	.	.	136	.	.	.	16	.	1	1			.
E. morrisbyi (ns)								ns			ns			
n	.		.	.	.	.	.	1			2	.	.	
E. perriniana (ns)				ns				ns						
n	.	.	.	0		.	.	7	.	.	.	.		.
E. rubida			ns	-	-			+			ns		ns	
n	.	.	2	49	0	.	.	63		.	1		0	.
E. urnigera (ns)				ns				ns						
n	.	.	.	25	.	.	.	3		.	.	.	.	
E. viminalis	-	+	-	+	-	+	-	+	+	-	+	ns	ns	ns
n	38	104	384	1455	326	34	42	496	58	27	161	8	59	127

Significant substrate preference or avoidance by each species was indicated by the sign of the residuals from the Chi-squared test of differences between the frequency distributions for presence and absence records, in different categories of geological type (Table 3.11). For example, the hypothesis of substrate fidelity was consistent for *E. pulchella* on Jurassic dolerite, and these patterns in the residuals are reversed for *E. tenuiramis* (Fig. 3.10). These two species, which occur across similar geographic and altitude ranges, appear to be partitioned across the available substrates. However, the substrate preferences suggested by the literature for *E. barberi*, *E. cordata* and *E. risdonii* (McEntee *et al.* 1994; Potts 1989; Wiltshire *et al.* 1992), are not confirmed by the current samples when these are constrained to the respective geographic and altitude domains (Table 3.11). The actual role of substrate is likely to be confounded with climatic factors which interact to influence the soil water availability, in addition to the effects of nutrient supply. Therefore, a suggestion of substrate fidelity which is independent of climate within a species' geographic and altitude range is inconclusive. For the purposes of predictive modelling, substrate factors may have a role in the definition of moisture availability, and therefore it is not appropriate to use geological type to further constrain sampling domains.

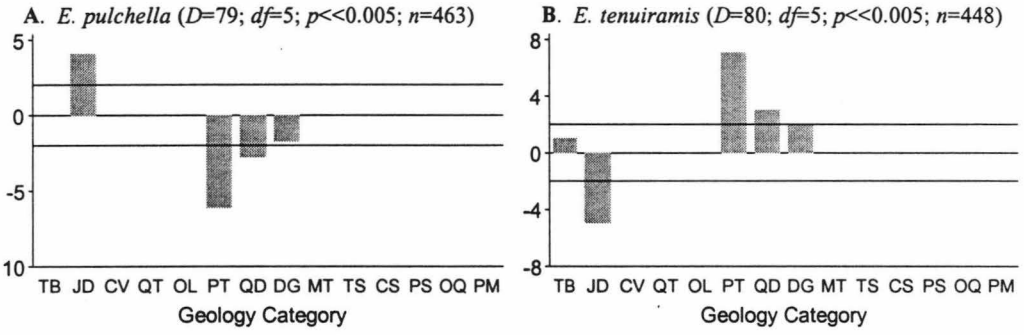


Figure 3.10 Relative substrate fidelity between *Eucalyptus pulchella* and *E. tenuiramis* as indicated by the Chi-square test between frequency of presence or absence observations within the respective geographic and altitude domains (See Table 3.11 for details and codes for rock type classes). Significant differences are indicated by the standardised residuals, $-2 > R_s > +2$, which fall outside the reference lines in each plot. The deviance (D), degrees of freedom (df), significance (p) and sample size (n) are also given.

3.3.5 Estimating minimum sampling requirements

Existing levels of environmental heterogeneity, summarised from separate classifications of biotic and abiotic attributes for the sample of observations in the ecological dataset are indicated by the frequencies in Tables 3.12 and 3.13 respectively. These summaries compare the variability of each set of presences with the corresponding set of absences for the geographic and altitude range of each species. The average number of replicate samples for each distinct type of environment is indicated by the ratio of samples to environments. An average of five samples for each type of environment was the arbitrarily selected minimum level of replication for representing ecological variability at the *ecoseries* to *ecosection* scale of classification (see Box 3.1). These criteria were used to assess sampling adequacy and to extrapolate this to an estimate of additional sampling requirements.

In the case of abiotic environmental heterogeneity (Table 3.13), only the presence records for *E. obliqua* comprise expected minimum levels of replication (average of five samples per type of environment), and for *E. amygdalina* there is an average of four replicate samples. Minimum levels of replication are more frequently achieved for the absence records, since these are usually far in excess of the number of presence records for each species. Extrapolation to five replicate samples in each case indicates that up to 2000 additional observations of presence and absence records for a species may be needed to adequately represent *ecoseries* scale ecological variability (Table 3.13). In the case of biotic environmental heterogeneity (Table 3.12), the fewer classified types of habitat lead to higher levels of replication and fewer requirements for additional sampling (up to 600 for presences or over 1000 for absences, depending upon the species).

Table 3.12 Biotic environmental heterogeneity of *Eucalyptus* species' occurrences (presences & absences) in the ecological dataset (see classification of biotic attributes in Table 2.3, Chapter 2). The average number of replicate samples for each distinct type of environment is indicated by the ratio of samples to environments (see empirical estimation method in Box 3.1). Additional minimum sampling [*n*(add)] was calculated from the difference between observed [*n*(obs)] and estimated [*n*(exp)] values.

	Presence					Absence				
Species	*n*(Biotic)	*n*(obs)	ratio	*n*(exp)	*n*(add)	*n*(Biotic)	*n*(obs)	ratio	*n*(exp)	*n*(add)
E. amygdalina	475	3991	8	2375	0	952	8321	9	4760	0
E. archeri	4	4	1	20	16	31	71	2	155	84
E. barberi	38	94	2	190	96	189	518	3	945	427
E. brookeriana	39	96	2	195	99	518	1745	3	2590	845
E coccifera	103	183	2	515	332	450	1921	4	2250	329
E cordata	45	53	1	225	172	268	734	3	1340	606
E. dalrympleana	218	672	3	1090	418	853	4500	5	4265	0
E. delegatensis	470	3147	7	2350	0	882	7533	9	4410	0
E. globulus	281	1097	4	1405	308	652	3980	6	3260	0
E. gunnii	38	70	2	190	120	451	1978	4	2255	277
E johnstonii	87	164	2	435	271	288	1192	4	1440	248
E. morrisbyi	4	4	1	20	16	22	29	1	110	81
E. nitida	128	339	3	640	301	497	4518	9	2485	0
E obliqua	575	8199	14	2875	0	858	5433	6	4290	0
E. ovata	334	1057	3	1670	613	992	9383	9	4960	0
E. pauciflora	136	254	2	680	426	655	3158	5	3275	117
E. perriniana	7	7	1	35	28	30	32	1	150	118
E. pulchella	182	515	3	910	395	607	3383	6	3035	0
E. aff. *radiata*	.	.	?	?	?	52	87	2	260	173
E. regnans	190	2467	13	950	0	758	6713	9	3790	0
E. risdonii	11	20	2	55	35	25	40	2	125	85
E. rodwayi	57	82	1	285	203	737	3924	5	3685	0
E. rubida	72	118	2	360	242	445	1172	3	2225	1053
E sieberi	85	182	2	425	243	326	856	3	1630	774
E. subcrenulata	9	13	1	45	32	171	397	2	855	458
E. tenuiramis	147	450	3	735	285	583	2945	5	2915	0
E. urnigera	19	30	2	95	65	197	450	2	985	535
E. vernicosa	.	.	?	?	?	76	115	2	380	265
E. viminalis	534	3321	6	2670	0	915	8466	9	4575	0

Table 3.13 Abiotic environmental heterogeneity of *Eucalyptus* species' occurrences (presences & absences) in the ecological dataset (see classification of abiotic attributes in Table 2.3, Chapter 2). The average number of replicate samples for each distinct type of environment is indicated by the ratio of samples to environments (see empirical estimation method in Box 3.1). Additional minimum sampling [*n*(add)] was calculated from the difference between observed [*n*(obs)] and estimated [*n*(exp)] values.

	Presence					Absence				
Species	*n*(Abiotic)	*n*(obs)	ratio	*n*(exp)	*n*(add)	*n*(Abiotic)	*n*(obs)	ratio	*n*(exp)	*n*(add)
E. amygdalina	1060	3991	4	5300	1309	1920	8321	4	9600	1279
E. archeri	4	4	1	20	16	42	71	2	210	139
E. barberi	33	94	3	165	71	152	518	3	760	242
E. brookeriana	60	96	2	300	204	555	1745	3	2775	1030
E. coccifera	95	183	2	475	292	723	1921	3	3615	1694
E. cordata	36	53	1	180	127	272	734	3	1360	626
E. dalrympleana	300	672	2	1500	828	1268	4500	4	6340	1840
E. delegatensis	1019	3147	3	5095	1948	1729	7533	4	8645	1112
E. globulus	445	1097	2	2225	1128	938	3980	4	4690	710
E. gunnii	44	70	2	220	150	708	1978	3	3540	1562
E. johnstonii	97	164	2	485	321	376	1192	3	1880	688
E. morrisbyi	3	4	1	15	11	16	29	2	80	51
E. nitida	189	339	2	945	606	1140	4518	4	5700	1182
E. obliqua	1754	8199	5	8770	571	1481	5433	4	7405	1972
E. ovata	318	1057	3	1590	533	1867	9383	5	9335	0
E. pauciflora	131	254	2	655	401	955	3158	3	4775	1617
E. perriniana	5	7	1	25	18	20	32	2	100	68
E. pulchella	233	515	2	1165	650	738	3383	5	3690	307
E. aff *radiata*	0	0	0	0	0	56	87	2	280	193
E. regnans	804	2467	3	4020	1553	1648	6713	4	8240	1527
E. risdonii	11	20	2	55	35	28	40	1	140	100
E. rodwayi	38	82	2	190	108	1110	3924	4	5550	1626
E. rubida	68	118	2	340	222	483	1172	2	2415	1243
E. sieberi	131	182	1	655	473	419	856	2	2095	1239
E. subcrenulata	13	13	1	65	52	229	397	2	1145	748
E. tenuiramis	212	450	2	1060	610	682	2945	4	3410	465
E. urnigera	21	30	1	105	75	175	450	3	875	425
E. vernicosa	0	0	0	0	0	76	115	2	380	265
E. viminalis	1048	3321	3	5240	1919	1879	8466	5	9395	929

The average biotic or abiotic environmental heterogeneity, for either presence or absence records of each species, were estimated for the sampled set of 100 km^2 cells from the observations in the ecological dataset (Table 3.14). These estimates were extrapolated into the unsampled cells of a species' sampling domain (Table 3.15). In each case, the relative levels of additional sampling for minimum replication of biotic or abiotic environmental heterogeneity were estimated. The combined estimate of additional sampling requirements for the occurrence (presence and absence) of each species, in the context of biotic or abiotic heterogeneity, provides an indication of the relative levels of ecological representativeness for the respective sampling domain (Table 3.16). Species were thus ranked, in Table 3.16, in general order of their representativeness within the ecological dataset.

Table 3.14 Representation in the sampled data domain. Relative levels of additional samples (presences & absences) needed to replicate biotic or abiotic environmental heterogeneity within the sampled cells of each species' geographic and altitude domains (based on data from Tables 3.4, 3.5, 3.12 & 3.13). A minimum sampling requirement [$\hat{n}$ (exp)] was empirically estimated as the product of the mean number of environments per 100 km^2 cell, the number of cells represented and a minimum of five observations (see Box 3.1). Additional sampling requirement [$\hat{n}$ (add)] was calculated as the difference between the observed and expected sample sizes.

	Presence					Absence				
		Abiotic		Biotic			Abiotic		Biotic	
Species	*n*(cell)	$\hat{n}$ (exp)	$\hat{n}$ (add)	$\hat{n}$ (exp)	$\hat{n}$ (add)	*n*(cell)	$\hat{n}$ (exp)	$\hat{n}$ (add)	$\hat{n}$ (exp)	$\hat{n}$ (add)
E. amygdalina	306	4590	599	3060	0	341	10230	1909	5115	0
E. archeri	2	20	16	20	16	12	240	169	180	109
E. barberi	12	180	86	180	86	17	765	247	935	417
E. brookeriana	28	280	184	140	44	62	2790	1045	2480	735
E. coccifera	58	580	397	580	397	97	3395	1474	2425	504
E. cordata	19	190	137	190	137	23	1380	646	1380	646
E. dalrympleana	117	1755	1083	1170	498	164	6560	2060	4100	0
E. delegatensis	235	4700	1553	2350	0	272	8160	627	4080	0
E. globulus	132	1980	883	1320	223	158	4740	760	3160	0
E. gunnii	33	165	95	165	95	78	3510	1532	2340	362
E. johnstonii	37	555	391	370	206	47	1880	688	1410	218
E. morrisbyi	2	20	16	20	16	2	80	51	110	81
E. nitida	86	860	521	430	91	123	5535	1017	2460	0
E. obliqua	331	8275	76	3310	0	329	8225	2792	4935	0
E. ovata	236	1180	123	1180	123	319	9570	187	4785	0
E. pauciflora	87	870	616	870	616	136	4760	1602	3400	242
E. perriniana	3	30	23	30	23	4	100	68	160	128
E. pulchella	78	1170	655	780	265	117	3510	127	2925	0
E. aff. *radiata*	0	?	?	?	?	6	270	183	270	183
E. regnans	127	3810	1343	635	0	162	8100	1387	4050	0
E. risdonii	3	60	40	60	40	4	140	100	120	80
E rodwayi	40	200	118	200	118	114	5700	1776	3420	0
E. rubida	38	380	262	380	262	74	2590	1418	2220	1048
E. sieberi	23	690	508	460	278	37	2035	1179	1665	809
E subcrenulata	3	60	47	45	32	27	1080	683	810	413
E. tenuiramis	70	1050	600	700	250	103	3605	660	3090	145
E. urnigera	9	90	60	90	60	25	875	425	1000	550
E. vernicosa	0	?	?	?	?	11	385	270	385	270
E. viminalis	316	4740	1419	3160	0	340	10200	1734	5100	0

Table 3.15 Representation in the unsampled data domain. Relative levels of additional samples (presences & absences) needed to replicate biotic or abiotic environmental heterogeneity within the unsampled cells of each species' geographic and altitude domains (based on data from Tables 3.4, 3.5, 3.12, 3.13 & 3.14). A minimum sampling requirement [$\hat{n}$ (add)] was empirically estimated as the product of the mean number of environments per represented cell, extrapolated to the number of unsampled cells [$\hat{n}$ (env)] for a minimum of five samples to adequately represent heterogeneity (see Box 3.1).

	Presence					Absence				
		Abiotic		Biotic			Abiotic		Biotic	
Species	*n*(cell)	$\hat{n}$ (env)	$\hat{n}$ (add)	$\hat{n}$ (env)	$\hat{n}$ (add)	*n*(cell)	$\hat{n}$ (env)	$\hat{n}$ (add)	$\hat{n}$ (env)	$\hat{n}$ (add)
E. amygdalina	112	336	1680	224	1120	77	462	2310	231	1155
E. archeri	18	36	180	36	180	8	32	160	24	120
E. barberi	5	15	75	15	75	0	0	0	0	0
E. brookeriana	52	104	520	52	260	18	162	810	144	720
E. coccifera	81	162	810	162	810	42	294	1470	210	1050
E. cordata	6	12	60	12	60	2	24	120	24	120
E. dalrympleana	59	177	885	118	590	12	96	480	60	300
E. delegatensis	115	460	2300	230	1150	78	468	2340	234	1170
E. globulus	79	237	1185	158	790	53	318	1590	212	1060
E. gunnii	68	68	340	68	340	23	207	1035	138	690
E. johnstonii	10	30	150	20	100	0	0	0	0	0
E. morrisbyi	0	0	0	0	0	0	0	0	0	0
E. nitida	206	412	2060	206	1030	69	1521	7605	676	3380
E. obliqua	107	535	2675	214	1070	09	545	2725	327	1635
E. ovata	176	176	880	176	880	93	558	2790	279	1395
E. pauciflora	82	164	820	164	820	33	231	1155	165	825
E. perriniana	1	2	10	2	10	0	0	0	0	0
E. pulchella	49	147	735	98	490	10	60	300	50	250
E. aff. *radiata*	7	?	?	?	?	1	9	45	9	45
E. regnans	44	264	1320	44	220	9	90	450	45	225
E. risdonii	1	4	20	4	20	0	0	0	0	0
E. rodwayi	95	95	475	95	475	21	210	1050	126	630
E. rubida	48	96	480	96	480	12	84	420	72	360
E. sieberi	14	84	420	56	280	0	0	0	0	0
E. subcrenulata	61	244	1220	183	915	37	296	1480	222	1110
E. tenuiramis	45	135	675	90	450	12	84	420	72	360
E. urnigera	20	40	200	40	200	4	28	140	32	160
E. vernicosa	75	?	?	?	?	64	448	2240	448	2240
E. viminalis	108	324	1620	216	1080	84	504	2520	252	1260

Table 3.16 Relative representativeness of environmental heterogeneity in the existing sample for each species (after Tables 3.14 & 3.15). Species are ranked by the ratio of extrapolated to observed sample sizes [$\hat{n}$ (add):*n*(plot)] for representation of biotic or abiotic environmental heterogeneity within the known geographic and altitude ranges of each *Eucalyptus* species (after Williams & Potts 1996).

	Abiotic						Biotic					
	Presence			Absence			Presence			Absence		
Species	$\hat{n}$ (add)	*n*(plot)	Ratio	$\hat{n}$ (add)	*n*(plot)	Ratio	$\hat{n}$ (add)	*n*(plot)	Ratio	$\hat{n}$ (add)	*n*(plot)	Ratio
Unrepresented:												
E. vernicosa	?	0	?	2510	115	21.8	?	0	?	2240	115	19.5
E. aff. *radiata*	?	0	?	228	87	2.6	?	0	?	45	87	0.5
E. subcrenulata	1267	3	97.5	2163	397	5.4	915	3	305.0	1110	397	2.8
E. archeri	196	2	49	329	71	4.6	180	2	90.0	120	71	1.7
Poorly Represented:												
E. urnigera	260	28	8.7	565	450	1.3	200	28	7.1	160	450	0.4
E. nitida	2581	384	7.6	8622	4518	1.9	1030	384	2.7	3380	4518	0.7
E. brookeriana	704	95	7.3	1855	1745	1.1	260	95	2.7	720	1745	0 4
E. rodwayi	593	82	7.2	2826	3924	0.7	475	82	5.8	630	3924	0.2
E. coccifera	1207	185	6.6	2944	1921	1.5	810	185	4.4	1050	1921	0.5
E. rubida	742	116	6.3	1838	1172	1.6	480	116	4.1	360	1172	0.3
E. gunnii	435	71	6.2	2567	1978	1.3	340	71	4.8	690	1978	0.3
E. pauciflora	1436	253	5.7	2757	3158	0 9	820	253	3.2	825	3158	0.3
E. sieberi	928	182	5.1	1179	856	1.4	280	182	1.5	0	856	0.0
E. perriniana	33	7	4.7	68	32	2.1	10	7	1.4	0	32	0 0
E. morrisbyi	16	4	4	51	29	1.8	0	4	0.0	0	29	0.0
Reasonably Represented:												
E. cordata	197	52	3.7	766	734	1.0	60	52	1.2	120	734	0.2
E. johnstonii	541	164	3.3	688	1192	0.6	100	164	0.6	0	1192	0.0
E risdonii	60	20	3	100	40	2.5	20	20	1.0	0	40	0.0
E. dalrympleana	1968	674	2.9	2540	4500	0.6	590	674	0 9	300	4500	0 1
E. tenuiramis	1275	447	2.8	1080	2945	0.4	450	447	1.0	360	2945	0.1
E. pulchella	1390	462	2.7	427	3383	0.1	490	462	1.1	250	3383	0.1
E. globulus	2068	1092	1.9	2350	3980	0.6	790	1092	0.7	1060	3980	0 3
E. barberi	161	94	1.7	247	518	0.5	75	94	0.8	0	518	0.0
E. delegatensis	3853	3141	1.2	2967	7533	0.4	1150	3141	0.4	1170	7533	0.2
E. regnans	2663	2463	1.1	1837	6713	0.3	220	2463	0.1	225	6713	0.0
E. ovata	1003	1055	0.9	2977	9383	0.3	880	1055	0.8	1395	9383	0.1
E. viminalis	3039	3318	0 9	4254	8466	0.5	1080	3318	0.3	1260	8466	0.1
Well Represented:												
E. amygdalina	2279	3991	0.6	4219	8321	0.5	1120	3991	0.3	1155	8321	0.1
E. obliqua	2751	8183	0.3	5517	5433	1.0	1070	8183	0.1	1635	5433	0 3

Widespread species such as *E. ovata*, which was sampled across 60% of its spatial range (Table 3.2), is quite well represented when compared with other species that occur in more variable habitat types (e.g. *E. globulus*). However, *E. obliqua* which is represented across three-quarters of its spatial extent, also exists across a wide range of habitats. Extrapolation of this ecological variability into the unsampled cells suggests that a relatively large number of samples may be needed to adequately represent the expected variability of habitats at the *ecoseries* to *ecosection* scale (Table 3.16). These trends in representativeness indicate that predictive analyses may not be appropriate for 15 of the species, given existing levels of sampling in the ecological dataset. Any analyses of species that are less than reasonably represented (Table 3.16) will require additional sampling, or constraints on the spatial extrapolation of predictions. Interpretation of the results of ecological analyses conducted for species which are reasonably represented may still be inconsistent with future analyses that are based on more comprehensive sampling. More

confidence in the repeatability of analyses for *E. amygdalina* and *E. obliqua*, which are well represented relative to the other species, is expected.

The ratio of biotic to abiotic environmental heterogeneity for a species' sample indicates potential differences in habitat specificity, and therefore differences in ecological processes, where this departs from the average for all species (0.54, in Table 3.17). For example, high levels of biotic to abiotic heterogeneity may be due to the influence of temporal disturbances on vegetation response, apparent in levels of habitat diversity (e.g. Houle 1994; Williams *et al.* 1994). Although low sampling frequencies and relatively low levels of representativeness confound the comparison in some cases, general differences in habitat between species can be inferred.

Table 3.17 Relative habitat differences between *Eucalyptus* species estimated from the ratio of biotic to abiotic heterogeneity (summarised from Tables 3.4, 3.13 & 3.14). Species are presented in increasing order of the Ratio (biotic/abiotic). Values for the ratio above the sample average of 0.54 (i.e. 1384 biotic vs. 2560 abiotic environments) indicate an accumulation of biotic heterogeneity at faster rates than the spatial occurrence of abiotic environments (not a ratio > 1 because the maximum number of biotic or abiotic environments is not the same).

Species	n(plots)	Range	Biotic	Abiotic	Ratio
E. rodwayi	82	40	57	38	1.5
E. perriniana	7	3	7	5	1.4
E. morrisbyi	4	2	4	3	1.3
E. cordata	53	19	45	36	1.3
E. barberi	94	12	38	33	1.2
E. coccifera	183	58	103	95	1.1
E. rubida	118	38	72	68	1 1
E. ovata	1057	236	334	318	1.1
E. pauciflora	254	87	136	131	1.0
E. archeri	4	2	4	4	1.0
E. risdonii	20	3	11	11	1.0
E. urnigera	30	9	19	21	0.9
E. johnstonii	164	37	87	97	0.9
E. gunnii	70	33	38	44	0 9
E. pulchella	515	78	182	233	0.8
E. dalrympleana	672	117	218	300	0.7
E. tenuiramis	450	70	147	212	0 7
E. subcrenulata	13	3	9	13	0.7
E. nitida	339	86	128	189	0.7
E. brookeriana	96	28	39	60	0.7
E. sieberi	182	23	85	131	0.6
E. globulus	1097	132	281	445	0.6
E. viminalis	3321	316	534	1048	0.5
E. delegatensis	3147	235	470	1019	0.5
E. amygdalina	3991	306	475	1060	0.4
E. obliqua	8199	1331	575	1754	0.3
E. regnans	2467	1127	190	804	0.2

The commoner species in the ecological dataset have a lower ratio of biotic to abiotic heterogeneity compared with the average for all *Eucalyptus* species (Table 3.17). This comparison indicates that, for these species, the sampled levels of biotic heterogeneity may be reasonably well explained by spatial attributes of the abiotic environment. The habitat of these species may be relatively uniform in time, on average influenced by longer-term factors of heterogeneity such as climate, rather than short-term disturbances that may lead to increased

diversity in the floristic and structural composition of the vegetation. *Eucalyptus regnans* has the lowest ratio of biotic to abiotic heterogeneity. This species probably occupies one of the most temporally stable habitats of all eucalypts, being characteristically warm, moist and fertile (e.g. Ashton 1981). In this equable environment, it achieves the tallest stature of all angiosperms (Boland *et al.* 1985); but in the absence of catastrophic disturbance by fire, which ensure episodic regeneration at 300 to 500 year intervals from canopy-stored seed, this habitat becomes occupied by rainforest (Gilbert 1959; Jackson 1968).

The ratio of biotic to abiotic heterogeneity increases toward the subalpine and dry sclerophyll habitats (Table 3.17). For example, a relatively high level of biotic variability (Ratio > 1.0) is apparent for the lowland dry sclerophyll species *E. pulchella* and *E. tenuiramis*, but even higher levels are apparent for the subalpine species *E. rodwayi*, *E. coccifera* and *E. pauciflora*. Short-term disturbance factors such as frequent, low intensity fires and drought in the lowlands, and climate extremes, such as wind, frost, snow and ice-glazing in the highlands, especially during the growing seasons, may act to increase habitat heterogeneity. The relatively high levels of spatio-temporal heterogeneity that may be expected from subalpine forest habitats may explain the comparatively high levels of floristic and structural diversity observed in the ecological dataset for species that are typical of such highland environments.

Sampling levels which realistically reflect the degree of ecological variability within a species' geographic and altitude range are the objective of this analysis. In this respect, it would appear that species of the dry sclerophyll and subalpine habitats which experience frequent disturbances of different types (e.g. frost, drought, fire) may need to be represented at higher levels than the lowland wet forest species which experience infrequent disturbances of a recurrent type (e.g. fire). However the extrapolation of these factors across the spatial domains for each species will require a great deal more information than is currently collated.

3.4 Discussion

The precision with which plant species' distributions can be predicted is enhanced when absence information is also taken into account (e.g. Austin & Cunningham 1981). The correlation patterns in a linear regression model between the presence and absence records and associated explanatory variables, with an appropriate transformation (e.g. logistic link function, McCullagh & Nelder 1989), defines the environmental response of a species. Apart from prediction, the shape of this response is the basis for interpreting the landscape patterns of vegetation and environment in terms of ecological processes. Therefore, absence records beyond the environmental limits of a species, or bias in the sample of observations, can distort this response function and lead to spurious predictions that confound ecological interpretation (Austin 1979; Austin & Meyers 1996). Depending on the study context, the potential or known distribution of a species is suitable for defining the sampling domain of presence and absence records, and for subsequently assessing environmental representativeness. However, the predictions of species'

distributions will ultimately be limited by the resolution of the data used for analysis (Norton & Williams 1992).

For these purposes, a sampling domain for *Eucalyptus* species in Tasmania was defined from their known geographic and altitude ranges (Williams & Potts 1996). These known distributions provide a repeatable and systematic method for obtaining the appropriate subset of presence/absence information from a compiled set of ecological data. A broad classification scale (100 km^2 cells and 100 m altitude classes) avoids the problem of a large number of absence records beyond the species' environmental range, but also ensures that the limits to distribution are well defined. The geographic and altitude ranges represent major indirect gradients of the environment that are strongly correlated with the climate and substrate factors that are more directly related to plant physiological responses (Austin & Smith 1989). The landscape definition of a sampling domain was selected for its simplicity in application and because it links pattern to process. Subsequent statistical modelling of species' distributions assumes that the sample data are representative of the landscape patterns of inter-relationships between the plant response and its habitat. However, this is not necessarily the case for compiled ecological data. The potential distribution of a species therefore has a two-fold application in predictive modelling — it provides the basis for defining the sampling domain for ecological data and a mechanism for assessing sampling adequacy.

The most efficient means of assessing representativeness within a sampling domain is by comparison with a theoretical sampling distribution from a designed biophysical survey (e.g. Margules & Austin 1994). However, landscape information can rarely match the scale at which observations are recorded, or at which predictions are intended. Therefore, several different approaches to assessing sampling adequacy may be needed. In the previous chapter (Chapter 2), the overall sampling adequacy of eucalypt forest habitat was assessed from surrogate estimates for ecological variability and from rarefaction curves of environmental heterogeneity. However, the appropriately-referenced attributes to link these information sources with the 100 km^2 grid-cell scales of individual species' distributions were not available to this study. Rarefaction methods (e.g. Colwell & Coddington 1994) could also be applied to assessing the sampling adequacy of individual species' distributions, but was computationally too intense to be considered here for each species' case. Nevertheless, rarefaction would be a suitable exploratory data analysis tool for assessing sampling adequacy on a case by case basis, although the method also extrapolates sampling bias (Colwell & Coddington 1994). For simplicity, and to obtain an indication of the trends in sampling adequacy compared between individual species, a numerical approach to extrapolating environmental heterogeneity from the sampled to the unsampled domains was adopted.

The potential geographic and altitude ranges for each species were therefore used to assess spatial sampling adequacy as a surrogate for environmental representativeness. However, comparisons with biotic contexts, in addition to abiotic patterns of sampling adequacy, were

needed to reflect the dominant processes influencing plant distribution patterns. Plant community classifications (e.g. Gillison & Anderson 1981) are a record of biotic variability and provide a suitable context for comparing the sampling adequacy of compiled data. In Tasmania, inventory surveys of forest vegetation have resulted in the classification of communities within wet (Jarman *et al.* 1984; Kirkpatrick *et al.* 1988a) and dry (Duncan & Brown 1985) forest types based upon the presence or dominance of indicator tree species and understorey characteristics. *Eucalyptus* species tend to dominate these communities (Kirkpatrick & Dickinson 1984). These classifications therefore provided a broad-scale indication of the variation in community structure and floristic composition that may be expected in eucalypt forest types, defined by the dominant species (Table 3.2).

The community classifications for eucalypt forest types in Tasmania (Kirkpatrick *et al.* 1988a, 1995; Duncan & Brown 1985) enabled species to be ranked according to their representation within the ecological dataset for the classification scale varying between *ecosection* and *ecoregion* (Table 3.3). The results from the spatial analysis indicated that fewer than half of the Tasmanian *Eucalyptus* species (44%) were represented across two-thirds or more of their potential geographic and altitude ranges (Table 3.4). With the inclusion of absence records, this level of spatial representation was achieved for 63% of species. Levels of replication across these spatial units indicated that some species' distributions were represented at a scale which approximated *ecosection* to *ecodistrict* (Table 3.6). However, when *ecoseries* scale environmental heterogeneity was extrapolated from the sampled to the unsampled ranges for each species, the expected patterns of ecological variability appear to be at least two to five times greater than current levels (Table 3.16). Representation of species' occurrences at the *ecoseries* scale of classification therefore requires a great deal more data than currently collated, but enabled species to be ranked according to relative levels of representativeness and sampling adequacy (Table 3.17). The combined results of these analyses indicate what types of restrictions to place on subsequent predictive modelling of species' distributions, given the existing set of compiled ecological data (Box 3.2).

Predictive analyses for the *ecosection* to *ecodistrict* scale will be possible for a subset of species in the existing set of ecological data (Box 3.2). Consistent with the assessment of sampling adequacy within eucalypt forest habitat units (Chapter 2), *Eucalyptus* species which predominate in Western, Midland and Highland regions are inadequately sampled. The western highland species, in particular, are unsampled or under-represented (*E. vernicosa* and *E. subcrenulata*). *Eucalyptus nitida*, which is widespread in lowland and highland habitats of western regions, is also poorly represented. With the advantage of hindsight, however, it appears that the published account of scrub and moorland communities which include *E. nitida* as an emergent (Jarman *et al.* 1988) should have been included in the earlier analysis of eucalypt forest community variability. Other widespread species with occurrences that are inadequately represented (e.g. *E. coccifera*, *E. urnigera*, *E. archeri*, *E. gunnii* and *E. rodwayi*) may reflect sampling bias

towards lowland habitats. Under-sampling among these species may be due to access constraints arising because either (i) their distributions are largely in remote western and highland habitats, or (ii) their main occurrences are on large private properties in the Midland regions.

Box 3.2 Relative suitability of *Eucalyptus* species for predictive modelling, with respect to occurrences in the ecological dataset.

Rare species that are unsampled or inadequately sampled for predictive modelling:
E. aft. *Radiata*
E morrisbyi
E. perriniana
E. risdonii

Western, Midland or Highland species that are inadequately sampled:
E. archeri
E. coccifera
E. gunnii
E. nitida
E. rodwayi
E. subcrenulata
E. urnigera
E vernicosa

Regionally-restricted species that are poorly represented:
E. rubida
E. sieberi
E. pauciflora

Regionally-restricted species that are reasonably represented:
E barberi
E. cordata
E. globulus
E. johnstonii
E. pulchella
E. tenuiramis

Widespread species that are reasonably represented across part of their range — suited to regional analyses only:
E brookeriana
E. ovata

Widespread species that are reasonably represented:
E dalrympleana
E delegatensis
E. regnans
E. viminalis

Widespread species that are well represented:
E. amygdalina
E obliqua

The sampling bias is also accentuated by the main data sources which were the continuous forest inventory (CFI) and the dry sclerophyll forest inventory (Table 3.1). The CFI data comprises the forest types with the most productive potential directing an environmental bias toward moist, lowland habitats throughout northern, eastern and southern regions of Tasmania. The dry sclerophyll data complement these productive forest types, but accentuates the bias toward lowland habitats and eastern regions (Duncan & Brown 1985). As a result, the *Eucalyptus* species that are well-represented are those which are of most interest for native forest management (e.g. *E. obliqua*, *E. regnans*, *E. delegatensis* and *E. globulus*). The rarer species, and species of poorer growth-form are less well represented, or not at all.

Some *Eucalyptus* species which are well-represented across part of their ranges could be adequately predicted for a specific set of environments by further constraining the data to a

specific spatial domain (Box 3.2). For example, *E. brookeriana*, which has two disjunct population centres in western and eastern regions (compare Fig. 3.8 with Williams & Potts 1996, p.55), may be adequately represented for predictive analyses based on the eastern population range. Similarly, occurrences of *E. ovata* may be considered for predictive analyses by excluding the unrepresented western populations from consideration (e.g. compare Fig. 3.4 with Williams & Potts 1996, p. 85). Even though some species may be inadequately represented in the ecological dataset, their presences may still contribute absence information for other species. Therefore, ecological factors other than species' distributions may also be of interest for analysis and prediction. For example, other biotic or abiotic classifications (e.g. biogeographic regions, forest type, taxonomic grouping) could be used to distinguish a sampling domain, depending upon the ecological relationships of interest for analysis.

Analyses of the distribution patterns of rare or uncommon species may require special considerations (e.g. Prober 1992). Rare *Eucalyptus* species (e.g. *E. morrisbyi*, *E. perriniana*, *E.* aff. *Radiata*, *E. risdonii*) are known from relatively few locations in Tasmania (e.g. see Table 3 in Williams & Potts 1996, pp. 124) and analyses based on occurrences (presence/absence) in forest quadrats (0.1 to 0.3 ha) may not be precise enough to distinguish their ecological relationships (Box 3.3). In addition, accurate predictions of either presence or absence for these species may be important when dealing with questions of land use (e.g. Blakesley & McDonald 1989). More specific measures of performance, such as abundance or biomass, and, possibly, more intensive sampling, may be needed for adequate prediction of their distribution patterns. However, for the uncommon species with broader distributions (e.g. *E. archeri*, *E. barberi*, *E. cordata*, *E. sieberi and E. urnigera*), presence and absence information may provide adequate resolution for prediction, but particular attention would need to be paid to environmental representativeness (Box 3.3). A bias toward absence information due to the regional abundance of commoner species may also have accumulated in the dataset (e.g. Table 3.10). The analysis of rare species may therefore require additional constraints on their sampling domain (e.g. Table 3.11), and an assessment of representativeness at a finer scale than 100 km^2 cells.

The inherent tendency for hybridisation and intergradation within *Eucalyptus* subgenera (Ellis *et al.* 1991), especially among closely related species (e.g. Duncan 1989a; updated in Williams & Potts 1996, p. 41), frequently results in problems of taxonomic classification. Competition or facilitation between species may also influence their distribution patterns. Some of these anomalies become apparent in the course of exploratory data analyses (e.g. Fig. 3.7, 3.9 & 3.10). Therefore, predictive modelling of some species may not be appropriate without the context of related or co-occurring species. For example, clinal species such as the yellow gums (e.g. Potts & Jackson 1986) and the white gums (e.g. Phillips & Reid 1980) can result in some problems of misidentification by field surveyors (see discussion for these species in Williams & Potts 1996, pp. 75, 110, 118). Compiled ecological data will inevitably include misidentifications and other sources of error. Therefore, knowledge of the general ecological and taxonomic relationships

between species can guide these exploratory analyses, and assist in the design of sampling domains and data requirements for predictive modelling (Box 3.3).

Box 3.3 Summary of general ecological and taxonomic relationships between *Eucalyptus* species which may influence exploratory analyses and the design of sampling domains and data requirements for predictive modelling.

1. Localised species, small geographic ranges
E. aff. *Radiata*
E. morrisbyi
E. perriniana
E. risdonii

Notes: Intensive sampling of specific performance attributes such as abundance or biomass, may be required in addition to absences, and presences should be representative and well replicated at local scales (e.g. 2 to 5 km^2 cells and 20 to 50 m altitude classes) throughout the respective geographic and altitude range.

2. Localised species, moderate geographic ranges
E. archeri
E. barberi
E. cordata
E. sieberi
E. urnigera

Notes: Extensive sampling of presence (or relative performance) and absence that is well replicated throughout the geographic and altitude range may be adequate for predictive analyses.

3. Regionally or environmentally restricted species
E. brookeriana
E coccifera
E. gunnii
E. johnstonii
E. pauciflora
E. pulchella
E. regnans
E. rubida
E. rodwayi
E. subcrenulata
E. tenuiramis
E. vernicosa

Notes: Extensive sampling of occurrences (presence and absence) that are reasonably replicated and representative of the geographic and altitude range may be adequate for predictive analyses.

4. Regionally or environmentally widespread species
E. amygdalina
E. dalrympleana
E. globulus
E. nitida
E. obliqua
E ovata
E. viminalis

Notes: Extensive sampling of occurrences (presence and absence) that are reasonably replicated and generally representative of the geographic and altitude range may be adequate for predictive analyses.

5. Groups of clinal or regionally intergrading species (see summaries in Williams & Potts 1996)
White gum cline: *E. viminalis*, *E. dalrympleana*, *E. rubida*
Yellow gum cline: *E. johnstonii*, *E subcrenulata*, *E. vernicosa*
Black gum intergrades: *E. brookeriana*, *E. ovata* in the south-west, west and north-west
Cider gum intergrades: *E gunnii*, *E. archeri*, *E. urnigera* in eastern Central Highlands
Wide-spread intergrading peppermints: *E. nitida*, *E. amygdalina* in the north-west
South-eastern intergrading peppermints: *E. amygdalina*, *E. pulchella*, *E. tenuiramis*
Highland intergrading peppermints: *E. coccifera*, *E. nitida* in east-west overlap zones
Wide-spread intergrading ashes: *E. obliqua*, *E. delegatensis* in mid-altitude ranges

Notes: Analyses of these species may need to be considered in the context of associated clines or intergrades, depending upon the species and the geographic region.

3.5 Conclusions

Species' geographic and altitude ranges provided a suitable context for delineating the sampling domain for predictive analyses based on presence and absence data, as well as the basis for assessing the representativeness of these data. This method of defining a sampling domain assumes that stratified random sampling within a species' potential geographic and altitude ranges will represent the pattern of ecological relationships that exist. Various methods of assessing sampling bias and representativeness of compiled data were used to distinguish between species which were suitable for further analysis and those which were inadequately sampled.

The resolution of the ecological dataset was estimated to be in the range of *ecosection* to *ecodistrict*, depending upon the species and the geographic region of analysis. This is consistent with the scale at which eucalypt forest habitat had been sampled (Chapter 2). Therefore, the collation of ecological information that is consistent with a stratification of key landscape factors, is also likely to yield data that is representative of species' distribution patterns. But the natural weighting toward common species in the landscape may require sampling for rare and uncommon species at a higher resolution than currently exists.

This assessment of sampling adequacy, and particularly the relative balance of presence and absence records for individual species' distributions, is an additional check on the statistical assumption of representativeness. Few presence cells relative to a large number of absence cells (and vice versa) were also potentially indicative of sampling bias that may lead to inaccurate predictions. The attributes used to assess representativeness were selected to provide a simple overview of sampling trends for each species, and depended on the availability of information that could provide a context for comparison with the range of ecological patterns and processes which exist in the landscape. More precise sources of information about the landscape, that could be matched with observations from the ecological dataset and with the potential distributions of individual species, would improve the accuracy of this assessment of sampling adequacy. The geographic and altitude ranges thus defined the spatial or environmental domain for sampling, as well as the domain in which predictions can be confidently interpolated.

Having broadly defined the potential for analysis of individual species' distributions, the physiological relevance of particular environmental gradients for predicting patterns of response needs to be considered. Questions of appropriate gradient definition and their application to predictive modelling are considered over the next three chapters.

4. Soil water supply: how well does a resource gradient estimated from limited site information predict species' distributions?

4.1 Introduction

The previous two chapters considered the sampling adequacy of a compiled set of ecological data intended for predictive analyses of *Eucalyptus* species' distributions in Tasmania. If the analysis of such inventory data are to provide ecological insight, the environmental gradients need to be as proximal as practical to plant physiological processes (e.g. Austin & Smith 1989; Austin & Gaywood 1994). However, the site information associated with field observations of plant performance is typically limited in the detail needed to distinguish local scale habitat variability (Walker & Langridge 1996). For example, field measures of environment may be limited to broad descriptions of parent rock type and topography (e.g. Austin *et al.* 1990; Mackey 1993a, b; Margules & Austin 1994). These minimum attribute data sets can then be used to generate other, more proximal, environmental variables by combining their location information with physical process models for climate or substrate, and digital elevation models of terrain (e.g. climate — McMahon *et al.* 1996; substrate — Moore *et al.* 1993a; Odeh *et al.* 1994, 1995; Mitasova *et al.* 1996).

Individual plant responses are only indirectly related to the average site estimates for broad-scale variation in climate and parent rock type, that are the commonly available levels of site information with inventory data. However, these average indicators of local site conditions may be reasonably correlated with the relative availability of resources at a site, with respect to the scale of response by the population of the species, rather than by the individual plant. Therefore, for the purpose of species distribution modelling with presence/absence response information, climate, terrain and substrate indices may be considered as relatively direct environmental gradients (*sensu* Austin & Smith 1989) with respect to their effect on population dynamic processes (e.g. overall demography, fecundity, biomass, stand dominance). For example, rainfall regimes determine the pattern of replenishment of soil water, the soil environment determines the storage capacity of the substrate, the terrain influences rainfall run-on/run-off characteristics and temperature regimes contribute to the evaporative demands of the atmosphere. These factors ultimately influence plant response processes of stress tolerance and biomass acquisition.

Climate and geological type provide the basic information necessary for distinguishing broad-scale site differences in vegetation habitat type (e.g. Haxeltine *et al.* 1996), and so with some assumptions about the basic conditions of the soil environment, this information could be reformulated as a gradient in soil water supply for clarifying species' response patterns in

predictive models. The physical relationships between the soil, the terrain and the atmosphere that influence water storage and runoff in vegetated landscapes have been extensively studied (e.g. Hatton *et al.* 1993; Moore *et al.* 1993; Pierce *et al.* 1993; Bouten 1995; Abdulla *et al.* 1996; Duan *et al.* 1996). These relationships have been applied to predictions of plant or crop growth responses which are highly species, scale and site specific, requiring detailed physiological information (e.g. Running & Coughlan 1988; Calder 1992; Barataud *et al.* 1995; Granier & Breda 1996; Jetten 1996; Thornley 1996). However, the limited data available with vegetation inventory surveys, reduces the precision with which site water relations can be estimated. The simplest models which may be applicable are based around the idea of the soil as a tipping-bucket (e.g. Fitzpatrick & Nix 1970; Keig & McAlpine 1974; Nix 1981; Porteous *et al.* 1994; Hobbs *et al.* 1994; Walker & Langridge 1996). Soil conditions within the root-zone of the plant (e.g. structure, texture, depth and rockiness) determine the size of the bucket used in a model of water balance. The replenishment of water to this rhizosphere follows the seasonal regimes of rainfall and evaporation, but deep water reservoirs and catchment drainage patterns can also contribute to plant water supplies (e.g. Dawson 1993, 1996; Zheng *et al.* 1996).

Plants require a relatively continuous supply of water as a consequence of gas-exchange during photosynthesis (Chapin *et al.* 1987; Schulze *et al.* 1987; Meinzer 1993; Losch & Schulze 1995). Stomatal conductance is therefore a measure of water loss and carbon gain, and depends on environmental factors such as temperature, soil moisture and carbon-dioxide concentrations, as well as atmospheric humidity levels and the aerodynamic effects of wind (Dewar 1995; Monteith 1995; Woodward *et al.* 1995). Plant water use can subsequently be determined by two major resistances: between the soil and the plant, and between the plant and the atmosphere. If we assume that plants fully utilise the soil volume, then the major component of the soil-to-plant resistance is the hydraulic conductivity, which is strongly influenced by soil water potential (e.g. Baratoud *et al.* 1995; Breda *et al.* 1995). Plants regulate stomatal conductance in response to root water potential (in addition to other factors such as the potential gradient across the root surface, the root surface area, and vapour pressure deficits), and all plants show a decline in stomatal conductance with declining soil water potential (e.g. Bates & Hall 1981; Schulze *et al.* 1987; Chapin *et al.* 1993; Castell & Terradas 1995; Triboulet *et al.* 1996). The actual shape of this response differs between species and between genotypes within species; but, for the purpose of estimating site water balance of a forest stand or an agricultural crop, it is generally referred to as a crop factor (e.g. McMurtrie *et al.* 1990; Whitehead & Kelliher 1991).

As a first approximation then, the long-term water use of a forest stand is directly proportional to soil water potential. The same relative water content gives rise to different soil water potentials in soils of different texture (e.g. Petersen *et al.* 1996; Cresswell & Paydar 1996). Therefore, a crop factor could be defined from the relationship between soil water content and soil water potential for a soil of a given texture. Such characteristic relationships as soil texture types, can be derived

from either simple experimental evaluation or approximated from text-book descriptions (e.g. Leeper 1964; Taylor & Ashcroft 1972; Murtha 1988).

There are two thresholds in plant response with declining soil water potential. At high soil water potentials, plants are unrestricted in their water use; but at very low soil water potentials, plants are no longer able to passively absorb water and become drought-stressed. Drought-stress in broad-leaved species is apparent as wilt, but some species have developed physiological and morphological adaptations, such as sclerophylly, to ameliorate the damaging effects of drying conditions (e.g. Morrow & Mooney 1974; de Lillis & Megrone 1994). Although plants vary in their ability to withdraw water from the soil under drought conditions, water potentials of about –1.5 MPa are generally referred to as wilting point (e.g. Shein & Pachepsky 1995; Dias-Filho & Dawson 1995; Sun *et al.* 1995; Myers & Talsma 1996). Wilting point is an important estimate of plant response because it reflects the available volume of water in the root-zone for a given set of soil conditions. The 'plant-available' water content of a soil is generally defined as the region between field capacity (equilibrium moisture conditions due to gravitational drainage following saturation) and wilting point (–0.01 to –1.5 MPa). Therefore a crop factor, derived from the characteristic water retention properties of a soil, would need to be estimated from the normalised relationships in the plant-available range of water content. Other workers have also used soil moisture extraction curves to represent the continuous decline in actual evapotranspiration with drying of the root-zone (e.g. Fitzpatrick & Nix 1970; Nix *et al.* 1977; Nix 1981; Hutchinson *et al.* 1982; Whitehead & Kelliher 1991; Nix *et al.* 1992; Walker & Langridge 1996; McMahon *et al.* 1996).

Soils with different texture can hold different volumes of available water and provide quite different types of habitat for plants (e.g. Gilliam *et al.* 1993). For example, one metre depth of clay soil could contain 130 mm of available water, but the same depth of coarse sand may only contain 80 mm of water. However, the characteristics of the soil environment (e.g. texture, structure) which determine water content, also influence the limiting rate at which water can be passively absorbed by the plant at different water contents. Clay soil can hold more water than sand, but the stronger matric potentials release water more gradually than is the case for sand. As a result, the interaction between climate and soil can lead to quite different environments in adjacent situations and result in different adaptive responses by plants (e.g. Lauenroth *et al.* 1994). For example, plants growing in clay soils experience a steady release of moisture between rainfall events, enabling continued, slow growth. This contrasts with adjacent plants growing in sand which may rapidly deplete available water between rainfall events, requiring more rapid responses to the onset of drought conditions. However, the soil environment has other tradeoffs for plants (e.g. Doran & Parkin 1994). Sandy soils can be more completely explored by plant roots, where as clay soils may be increasingly difficult to penetrate as they dry out. Clay soils may also be more difficult to wet up once dry, and intermittent rain may run off without penetrating to the root-zone. Alternatively rain falling on sandy soils may drain rapidly beyond

the root-zone. Therefore, sandy soils often have a higher water availability than clay soils, except during periods of drought. Nutrient status is another important difference between clay and sand soil texture types, also involving an interaction with the moisture regime, but these ecological aspects of site conditions will not be discussed further here.

Interactions between weather and the soil environment are therefore a key to the supply of water to a plant. The natural occurrence of a species reflects its adaptation to a particular water regime (e.g. Anderson *et al.* 1996). For example, three *Eucalyptus* species, *E. amygdalina*, *E. pulchella* and *E. tenuiramis*, from the series *Piperitae*, are frequently observed to occur on contrasting substrates within the same climatic regime of south-eastern Tasmania (Duncan & Brown 1985; see also Appendices 4 & 5). Differences in soil moisture regime could be attributed to the different substrates, and thereby contribute to an explanation of these species' responses (Davidson *et al.* 1981). In the absence of specific information about the stomatal response of species and their canopy surface area, information about the soil environment in the rhizosphere could provide a reasonable estimate of the potential water supply-rate to the plant with variation in soil moisture conditions.

Inventory data comprises information about the co-occurrence of species, in addition to the distribution of individual species. Since species' occurrences are influenced by the presence of other species, the potential for interaction between species which may be suggested by their overlapping distribution patterns is also of interest. The derivation of a water supply gradient needs to be able to address these comparative questions of ecology. Therefore, a single site estimate of soil water supply is needed, even though the actual experience of water stress by individual species within a site may differ. Water balance estimates based on the physiological response of one species, or an average physiological response for all *Eucalyptus* species, may not be an appropriate generalisation. For these reasons, a water balance model which adequately summarises potential environmental water relations from site variation in climate and soil, is considered a reasonable compromise for the purpose of comparative ecological study. Such a site estimate would not distinguish differences in the physiological response of individual species. However, where the purpose of a water balance model is to predict site-specific plantation growth of a forest tree species, alternative models based on the climate, soil and genetic parameters for the physiological response of the species will be more appropriate (e.g. Desänker *et al.* 1994; Korol *et al.* 1996; Luan *et al.* 1996; Thornley 1996).

Therefore the attributes required of a water balance model for predicting plant species' distributions include (i) an ability to utilise simple environmental information derived from climate and geological type, (ii) flexibility for incorporating additional climate and substrate information as this becomes available, and (iii) simplicity in estimation of moisture regimes. This chapter thus explores the potential for estimating a generic gradient in soil water supply from, simple sets of abiotic environmental information that would be useful for comparing the responses of different *Eucalyptus* species along the same moisture gradient.

To achieve these objectives, the simple tipping-bucket model of water balance, WATBAL, developed by CSIRO (Slatyer 1960; Fitzpatrick & Arnold 1964; Fitzpatrick 1965; McAlpine 1970; Keig & McAlpine 1974) was initially adopted. Variation in evapotranspiration with soil water content was subsequently determined from the soil water potential, and different relationships were derived for different classes of soil texture. Since the inventory data contained no specific information about the soil environment, the textural properties of a forest soil were approximated from geological types in consultation with forest soil scientists. In generating this model, the environmental parameters of climate and water balance associated with the set of compiled ecological data for the occurrence of *Eucalyptus* species in Tasmania were demonstrated.

The performance of this water balance model was assessed by comparing daily estimates of soil moisture content with neutron probe measurements at two closed-canopy *E. globulus* plantation sites. A third set of data, consisting of the growth potential of *E. globulus* in plantation, was used to assess the differences in estimating average soil water relations from the soil environment compared with estimates based on genetic parameters for stomatal conductance and canopy area. These data also enabled the loss in precision when calculating site water balance with limited information about the substrate to be assessed. Finally, the potential gains in both the interpretability of results and the overall explanatory power in subsequent predictive modelling of species' distributions, were assessed with four *Eucalyptus* species from contrasting habitats in Tasmania. The modelled response of each species to the univariate or multivariate gradients of water supply were compared with their response to climate.

4.2 Method

4.2.1 Ecological dataset

An estimate of site water relations for 15 640 observations of the presence or absence of *Eucalyptus* species in Tasmania was the objective of this analysis. The representation of species and environments in these data were previously described in Chapters 2 and 3. Site locations (latitude, longitude and altitude) enabled long-term monthly averages of climate (rainfall, rain days, pan evaporation, minimum and maximum temperature, flat-surface and cloud-terrain adjusted solar radiation) to be estimated from the process model, ESOCLIM (H. A. Nix, J. R. Busby, M. F. Hutchinson & J. McMahon; see McMahon *et al.* 1996). Geological types were derived from field observations, with missing values matched with the geological categories mapped at 1:500 000 scale (Department of Mines 1976), but no other site information (e.g. soil depth, rockiness, drainage, terrain position) were consistently recorded between data sources. Site information for modelling water balance was therefore limited to climate and geological type. This is typical of compiled ecological data (e.g. Austin *et al.* 1990; Austin & Meyers 1996).

4.2.2 Developing a model of site water balance

The simple water balance model WATBAL (McAlpine 1970), was considered by Nix (1981) to be suited for use with long-term mean values of precipitation and evaporation. WATBAL defines the water demand for period n (e.g. days, weeks, months) in the year as a function of the soil, vegetation and climate conditions which adjust the potential evaporation to an estimate of the actual evapotranspiration (see description of the WATBAL model in Box 4.1). A daily step was selected to realistically allocate variation in the patterns of rainfall per month.

Box 4.1 Calculation steps in WATBAL, a simple tipping-bucket water balance model (after McAlpine 1970; Keig and McAlpine 1974).

The water demand for period, n of the year is defined by:

(1) $NDMD_n = AETCF_n * PETCF_n * EVAP_n$

$NDMD_n$ is the water demand for period n. $AETCF_n$ is the coefficient defining the relationship between actual evapotranspiration and potential evapotranspiration (e.g. specific soil/vegetation characteristics). $PETCF_n$ is the coefficient defining the relationship between potential evapotranspiration and potential evaporation (i.e. terrain adjusted wet surface pan evaporation). $EVAP_n$ is the empirical estimate of pan evaporation from maximum temperature and relative humidity data.

(2) where $AETCF_n$ is dependent on $[(NSTR_{n-1} + RAIN_n)/MAXST]\%$, the relative available soil water storage ($R^*_{\theta s}$). (i.e. $AETCF_n = f(R^*_{\theta s})$)

$NSTR_{n-1}$ is the soil moisture store of the period prior to that for which calculations are being made. $RAIN_n$ is the rainfall of the current period. MAXST is the available soil moisture (field capacity - wilting point) for the depth of the root zone and dependent upon the properties of the soil texture class and bulk density.

A simple stepped or continuous relationship between the actual evapotranspiration coefficient and relative available soil water storage needs to be determined, depending upon the specific soil/vegetation characteristics restricting evaporation below the wet surface potential. For example, for Canberra, a two-step relationship of actual evapotranspiration coefficient to soil moisture and rainfall may consist of a threshold at $R^*_{\theta s_n} = 50\%$, above which $AETCF_n = 1.0$ and below which $AETCF_n = 0.5$.

Then soil moisture storage for each period, n, is derived by:

(3) $NSTR_n = (NSTR_{n-1} + RAIN_n) - NDMD_n$

(4) but if: $NDMD_n > (NSTR_{n-1} + RAIN_n)$; then $NSTR_n = 0$

(5) and if $NSTR_n > MAXST$

then the amount by which it is greater is defined as water surplus ($SPLS_n$) to give an indication of drainage, runoff and deep percolation

(6) $SPLS_n = NSTR_n - MAXST$

(7) and: $NSTR_n = MAXST$

(8) Actual evapotranspiration (AET_n) while not being directly calculated, will be equal to actual water demand

$AET_n = NDMD_n$ except where $NDMD_n > (NSTR_{n-1} + RAIN_n)$, in which case

(9) $AET_n = NSTR_{n-1} + RAIN_n$

These calculations estimate seasonal changes in the soil moisture regime at a site. The soil moisture regime could be used to define consecutive growth and drought periods. For example, (Fitzpatrick 1965) estimated growth periods where soil moisture storages were consecutively above 60% of field capacity. McApline (1970) suggested better results could be obtained if more steps were included in the functional relationship of AETCF with $R^*_{\theta s}$. The volume of water surplus could also be expressed relative to rainfall (e.g. Church *et al* 1995).

Quasi-daily rainfall patterns were estimated from average monthly rainfall and the number of rain-days per month. Each day of the month was assumed to have an equal probability of rainfall, although a clumped or Poisson distribution is more likely (e.g. Guenni *et al.* 1996). Since rain may fall on different days in each month, several years of continuous daily rainfall were estimated by generating random numbers from a uniform distribution (function RANUNI, SAS Institute Inc. 1990f). Monthly averages of daily pan evaporation were converted to quasi-daily values by interpolation. Differences between the actual and interpolated estimates, compared as monthly totals, were minor and may be attributed to the fact that the polynomial curve used in the interpolation was based upon the middle day of each month, whereas the actual monthly average may have been skewed.

A coefficient of potential evapotranspiration (PETCF), which adjusts pan evaporation to an estimate of potential evapotranspiration was derived from the effect of site orientation (slope and aspect) on insolation. In humid environments the ratio of potential evapotranspiration to pan evaporation is commonly set equal to one (Nix 1981). However at higher latitudes, where sun angles and day lengths have a marked seasonality, the influence of slope and aspect on site insolation and subsequent evapotranspiration cannot be ignored (e.g. Holland & Steyn 1975; Kirkpatrick & Nunez 1980; Nunez 1980, 1983). The seasonal and site variation in the ratio of flat to inclined surface solar radiation was used to terrain-adjust the quasi-daily estimates of pan evaporation toward potential evapotranspiration. This index was in part a surrogate for seasonal and site differences in vegetation development which are related to terrain through the solar radiation index. As a result, daily potential evapotranspiration may be more or less than the quasi-daily estimates for pan evaporation. The continuous daily estimates of potential evapotranspiration were matched with daily rainfall patterns in a model of water balance. Canopy interception of rainfall and subsequent wet-surface evaporation were not calculated, but potential evapotranspiration was assumed to occur every day, irrespective of a rain-day.

Both resistance to water uptake from the soil within the root-zone and stomatal conductance are assumed to be strongly correlated with soil water potential. A coefficient of actual evapotranspiration (AETCF), which adjusts potential evapotranspiration to an estimate of actual evapotranspiration, was therefore derived from the relationship between soil water potential and soil water content. This relationship is a characteristic property of soils and can be approximated from soil texture descriptions. Three soil texture classes, representing the expected clay content of soil developing over parent rock types in native forest habitats of Tasmania, were approximated by expert opinion from geological categories (1:500 000 scale geological Map of Tasmania, Department of Mines 1976) (B. Neilson personal communication, November 1994). These texture types, with intermediates defined in some cases, were related to the soil texture classes, sandy loam, silt/clay loam and clay, for estimating available (transpirable) water between field capacity and wilting point (Table 4.1).

Table 4.1 Mean available (transpirable) water retained between –0.01 and –1.5 MPa (field capacity to wilting point), for the soil texture classes and their intermediates (L, M/L, M, M/H, H: B. Neilson personal communication, November 1994), and moisture extraction classes (coarse sandy loam, clay loam, clay) for 100 cm profile depth fully explored by plant roots (after Taylor & Ashcroft 1972). The number of sites (*n*) in the ecological dataset comprising each texture class is also indicated.

Water extraction class	Soil Texture Class	Available Water (mm water per metre depth)	*n*
Coarse sandy loam (S)			
(L)	Coarse sandy loams	80	7292
(M/L)	Fine sandy loams	140	231
Silt/clay loam (L)			
(M)	Silt loams	170	2097
Clay (C)			
(H/M)	Silty clay loams	125	407
(H)	Clays	115	5613

The water retention curves that approximated these soil texture descriptions were derived from Taylor & Ashcroft (1972). Only those portions of the curves which represented the plant-available water fraction (field capacity to wilting point: –0.01 to –1.5 MPa) were transcribed to define these relationships. The coefficient of actual evapotranspiration (AETCF) was defined as a function of the fractional available soil water content (of the maximum available soil water storage) from the normalised relationship between soil water content ($R^*_{\theta s}$, in the plant-available range) and soil water potential (Ψ_s, MPa). A single-parameter equation for this normalised relationship (after Walker & Langridge 1996) was applied to each soil texture class (Fig. 4.1).

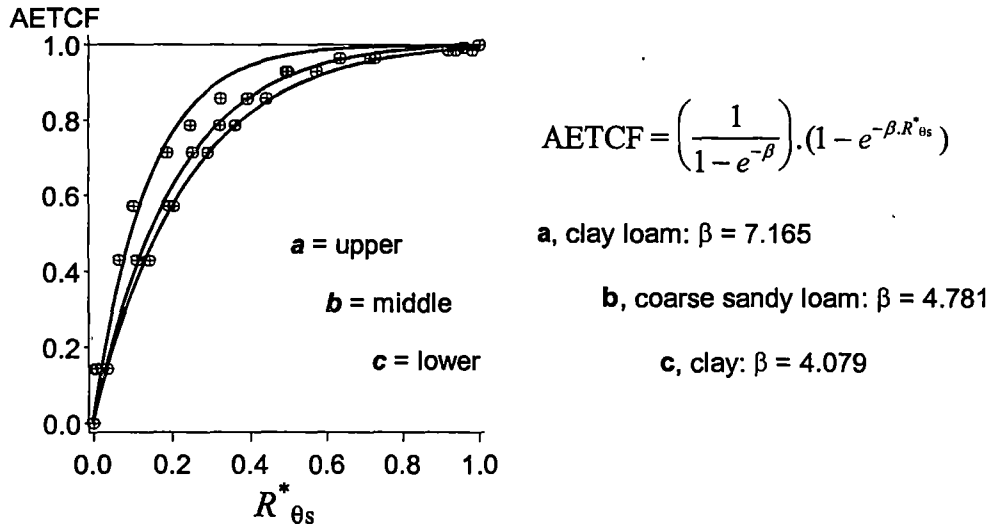


Figure 4.1 The continuous relationship between the actual evapotranspiration coefficient (AETCF) and the relative available soil water content ($R^*_{\theta s}$) defined from the water retention curves for three soil textures in the available range (field capacity to wilting point, –0.01 to –1.5 MPa); based on the single-parameter equation suggested by Mackenzie and co-workers (cited in Walker & Langridge 1996).

In the absence of other substrate detail with compiled ecological data, soil depth was defined as a constant of one metre for all sites, and rockiness was defined as zero for the purpose of estimating site water balance. One metre of soil depth is a reasonable overall value for

Tasmanian forest soils, with variation between a few tens of centimetres for peat soils on Precambrian quartzite and up to 2 m on Quaternary sands (after Grant *et al.* 1995). Using these formulations, the daily estimates of soil water content (NSTR) derived from the water balance model could be redefined as soil water potential by an exponential back-transformation as relevant to each soil texture class (Ψ_s = 1.5(*exponential function*)–1.5).

Although the normalised water retention curves for the coarse sandy loam and clay soils are similar (Fig. 4.1) the separate functions were retained, because it is anticipated that where actual site information for a particular water retention curve is available, these would be applied rather than a textbook approximation. The single parameter equation is a good fit to the coarse sandy loam and clay loam texture classes, but is a poor fit to the clay loam texture class. A three parameter equation of the form $\text{AETCF} = \alpha.e^{\beta/(\gamma+R^*_{\theta s})}$ (after Ratkowsky 1990) was found to provide a closer approximation to the normalised clay loam retention curve, but for reasons of parsimony and consistency it was not used here.

These normalised curves are all similar to the clay loams of Nix (1981; redefined as sandy loam in Walker & Langridge 1996), and do not reflect the wide differences between their three texture classes. In addition, the order of the texture classes in Fig. 4.1 for coarse sandy loam and clay loam, are reversed in comparison with those presented by Walker & Langridge (1996). It is possible that the normalised relationships used by Nix (1981; or Walker & Langridge 1996) may have been based upon the range in soil water contents between saturated to air dry, rather than the transpirable range (field capacity to wilting point), or they used a wider range of texture classes than selected here to represent Tasmanian forest soils.

4.2.3 Testing and validating the water balance model

Two additional sets of data were used to validate the approach taken for calculating water balance. The first dataset consists of a series of neutron probes measurements of soil moisture, taken between July 1992 and June 1993, at two contrasting *Eucalyptus globulus* plantation sites in Western Australia (Hingston *et al.* 1994). These soil moisture readings from sites at Mumbalup and Manjimup (Table 4.2) were compared with the soil moisture estimates from the water balance model. The complex, layered soil profiles, varying from fine sandy clay loams, sandy clay, clay loams to medium clay were not distinguished in this comparison. A mean soil texture class of the 'clay' type was selected from the three previously modelled soil water retention curves (Fig. 4.1). This comparison enabled the performance of the water balance model to be assessed on a daily step with actual rainfall events. However, only long-term averages for pan evaporation were available rather than actual site measurements of potential evapotranspiration.

Table 4.2 Average environmental conditions at Mumbalup and Manjimup *Eucalyptus globulus* plantation sites in Western Australia (after Hingston *et al.* 1994). Total annual rainfall (P_T) and evaporation (E_P), mean annual maximum (T_{max}) and minimum (T_{min}) temperatures, and mean annual net solar radiation (R_{net}) for the measurement period (365 days) during which soil moistures were recorded (July 1992 to June 1993). LAI refers to the plantation leaf area index, indicating the development of a closed canopy at each site (maximum potential transpiration rates). Long-term mean annual rainfall ($P_{T(MEAN)}$) and total available soil water storage (MAXST) are also shown.

Site	P_T (mm/year)	E_P (mm/year)	R_{net} (MJ m^{-2}/day)	T_{max} (°C/day)	T_{min} (°C/day)	LAI	$P_{T(MEAN)}$ (mm/year)	MAXST (mm)
Mumbalup	698	1647	11.1	22.4	7.8	4	950	250
Manjimup	983	1243	15.1	19.4	8.58	4.5	1050	204

The second dataset is a series of environmental parameters (Table 4.3) associated with 19 *E. globulus* plantation sites in Northern Tasmania (Laffan 1993, 1994; Osler *et al.* 1996). The extent to which water limits the growth of *E. globulus* across these sites was estimated from genetic parameters for stomatal conductance and canopy leaf area with the model PROMOD developed by Battaglia & Sands (1997). Comparison of PROMOD output with actual soil water fluxes indicates that it realistically simulates the general processes of site water relations (Battaglia & Sands 1997).

Table 4.3 Range in site information across 19 *Eucalyptus globulus* plantations in northern Tasmania (after Laffan 1993, 1994; Osler *et al.* 1996). MAXST is the maximum available water storage (integrating the effects of soil texture, soil depth and stoniness). Four soil texture classes are recorded, varying from uniform sands or loams or gradational soils with structured subsoils (14 sites) to duplex soils with structured clays or gradational soils with poorly structured subsoils (3 sites), structured uniform clays or duplex soils with massive clays (1 site), and massive uniform clays (1 site).

Site	Soil Depth (cm)	Stones (%)	Available water (cm/m depth)	MAXST (mm/site)	P_T (mm/yr)	E_P (mm/yr)	P_T–E_P (mm/yr)
Mean	100	16	139	121	1055	1052	2
Median	110	10	150	149	1034	1058	–42
Maximum	110	70	150	149	1457	1107	458
Minimum	38	10	80	27	858	971	–237

The water balance model of the soil environment was used to predict seasonal variation of actual evapotranspiration, soil water potential and soil water surplus for these 19 plantation sites. For this purpose, measures of the soil environment included texture, depth and rockiness. The ranking and magnitude of site droughtiness from PROMOD were compared (using a regression relationship) with the water balance variable for mean annual soil water potential. This comparison enabled the differences between the two methods of estimating a water supply gradient to be assessed.

The environmental parameters for *E. globulus* plantation sites in northern Tasmania also enabled assessment of the loss in precision when calculating site water balance with varying levels of information about the substrate. The relative importance of the additional complexity introduced

into the water balance model by substrate parameters for root-zone depth, structure (e.g. stone fraction) and site differences in texture were assessed by comparing the sequential improvement in the linear regression relation with the site index of water-stress limited growth from PROMOD.

Five models of site water relations were compared:

1. All substrate variation (depth, texture, stones as indicated by Laffan 1993 & Osler *et al.* 1996),
2. Soil depth constant at 100 cm, MAXST varying by substrate texture and stones.
3. Soil depth constant at 100 cm, no stones, MAXST varying by substrate texture only.
4. Soil depth varying, no stones, substrate texture as 'clay' retention curve (MAXST=130×depth, mm).
5. Soil depth constant at 100 cm, substrate textures unvarying ('clay' retention curve, MAXST=130 mm).

The absence of climate-substrate interactions were also considered in a comparison with the PROMOD index. The annual net climatic water balance (sum of the monthly differences between precipitation and evaporation) is a measure of potential environmental water supply which assumes no interaction with vegetation or substrate. The change in the degree of correlation (Pearson coefficients) between the two main independent components of water balance, soil water potential and soil water runoff, were compared for each of the five levels of information used to derive water scalars. These sequential analyses of the relative precision of a water supply gradient for various levels of limiting information were designed for comparison with information levels available with compiled ecological data.

4.2.4 Water balance variation in Tasmanian eucalypt forests

The level of site information available with the set of compiled ecological data for Tasmanian eucalypt forests was limited to climate with approximations of soil texture from parent rock type, but all other substrate conditions (e.g. soil depth, rockiness, horizon structure) were treated as constant (i.e. one metre depth, single soil layer, and no stones). The seasonal variation in water relations of these 15 640 native forest sites, as predicted from the water balance model of the soil environment, was demonstrated for soil water potential and soil water surplus. The influence of three soil texture classes in defining the relationship between these variables is also demonstrated. The degree of correlation (Pearson coefficient) between these two main components of water balance, soil water potential and soil water runoff, and climate factors (rainfall, evaporation, temperature, and net atmospheric water balance - precipitation minus evaporation) was compared. These correlations, based on the levels of information in compiled ecological data, was defined for comparison with the sequential analyses of water balance from the 19 *E. globulus* plantation sites in northern Tasmania.

4.2.4.1 Modelling performance of individual Eucalyptus *species from native forest stands*

Logistic regression models were used to distinguish the water-relations response of four *Eucalyptus* species, *E. regnans, E. obliqua, E. amygdalina* and *E. globulus*, from contrasting habitats in Tasmania (using PROC LOGISTIC, SAS Institute Inc. 1990d). Each of these species occurs as a canopy dominant and is generally found in lowland habitats across a wide range of

substrate types. They were chosen for the purpose of comparing responses to water relations, without large confounding effects due to temperature, light or nutrients. The method for selecting the relevant set of absence samples for each species from the ecological dataset was defined in Chapter 3.

The potential gains in both the interpretability of results and the overall explanatory power for subsequent predictive modelling of species' distributions from univariate gradients in soil water supply were compared with similar gradients in environmental water derived from climate. To assess the potential maximal model of the response for each species to water, the climate or water balance variables were fitted in the presence of a temperature variable (mean annual daily maximum or minimum). The shape of a species' response to each variable was described by polynomials. Polynomials higher than the fourth order were not considered, since these may be artifacts of the dataset, rather than indicative ecological responses (e.g. Austin & Meyers 1996). A quartic polynomial may represent a broad optimum, a strongly skewed response or two optima; consistent with a competition effect (Mueller-Dombois & Ellenberg 1974; Austin 1980, 1985, 1990). A cubic polynomial allowed a skewed response to be considered (Austin & Smith 1989, Austin 1990, Austin & Gaywood 1994), and a quadratic polynomial may represent a classic bell-shaped response (Gauch & Whittaker 1972). Models were derived by backward elimination of the highest order polynomial, until all remaining highest order polynomials for a variable were significant (after Nelder 1994).

4.3 Results

4.3.1 Testing and validating a water balance model

4.3.1.1 Western Australian soil moisture data

Comparison between the estimation of soil water content and actual measurements of soil moisture indicates that this simple model of water balance, based on the soil water retention function, realistically simulates the general processes of site water relations. Summary statistics for the observed and predicted daily soil water contents for two closed-canopy *Eucalyptus globulus* plantation sites are presented in Table 4.4. The magnitude of the relative differences between the observations for soil water content and predictions derived from the model of water balance were assessed for each comparable time for each site (Fig. 4.2). The smallest deviation (4% maximum vertical difference between the empirical cumulative distributions) occurred for the Manjimup site (Fig. 4.2a). Predictions of daily soil water status were also comparable (12% maximum vertical difference) for the Mumbalup site (Fig. 4.2b).

Table 4.4 Annual mean, median and standard errors for the observed (θ_s) and predicted ($\hat{\theta}_s$) estimates of soil water content over the measurement period (July 1992 to June 1993) for two *Eucalyptus globulus* plantation sites in Western Australia (after Hingston *et al.* 1994).

	Mean		Median		Standard Errors	
Site	θ_s (vvw)	$\hat{\theta}_s$ (vvw)	θ_s (vvw)	$\hat{\theta}_s$ (vvw)	θ_s (vvw)	$\hat{\theta}_s$ (vvw)
Mumbalup	117	94	73	54	25	5
Manjimup	105	102	110	103	20	4

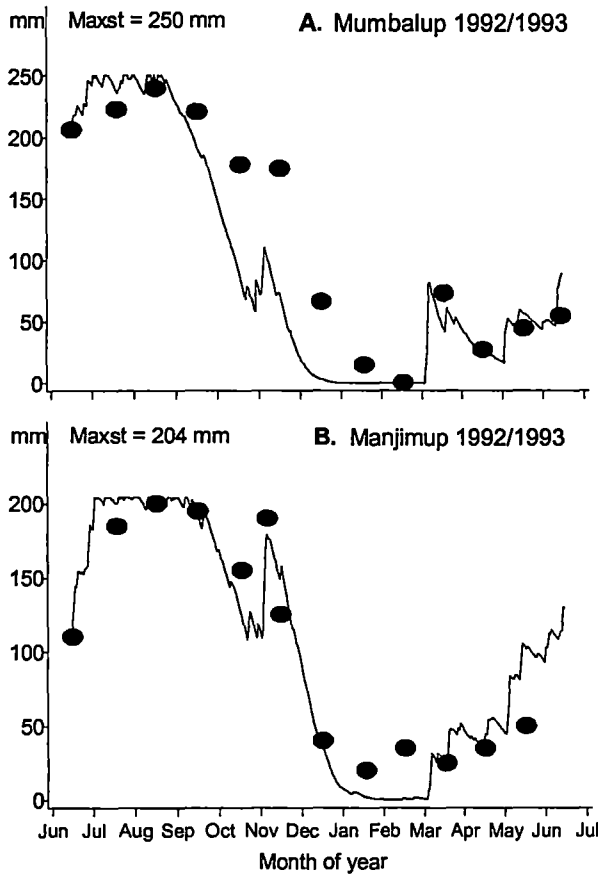


Figure 4.2 Comparison between observed (filled circles, measured June 1993 to March 1993) and predicted (daily trace) soil water contents for Mumbalup (**A**) and Manjimup plantations (**B**), established in 1988 (see summaries of environmental data in Table 4.2).

Daily evaporation was estimated from monthly mean values derived from ESOCLIM.

Daily rainfall was obtained from site measurements, and observed available soil water (mm) was assessed from monthly neutron probe readings (after Hingston *et al.* 1994).

The water balance was modelled with the water desorption curve for soils of clay texture (see Fig. 4.1).

The water balance model was further evaluated using several model performance measures (Table 4.5). All performance measures suggested that the model may increasingly over-estimate the reduction in soil water content as the mean annual rainfall at the site declines (*cf.* Table 4.2). That is, the *E. globulus* plantation was drying the soil at lower rates than the modelled drying curve indicated. Some error could also have arisen from assumptions about the soil environment (single layer, assumed texture, water relations approximated from text books), and from the use of long-term averages for pan evaporation (estimated through ESOCLIM) in the water balance modelling, rather than actual site measurements. However, the regression relationship between the observed and predicted soil water contents, matched for the 13 daily measurements, indicate that a reasonable estimate of average soil water content was made for each site ($r^2_{adj.}$~ 90%, 97%).

Table 4.5 Evaluating the performance of the water balance model. Comparison of predicted estimates of soil water content with 13 observations over a 12 month measurement period (July 1992 to June 1993) for two *Eucalyptus globulus* plantation sites in Western Australia (after Hingston *et al.* 1994). RMB is the relative mean bias, IOA is the index of agreement, NMAE is the normalised mean absolute error and the slope parameter β, its standard error (β stderr) and the adjusted r^2 are derived from the linear regression relationship $\theta_{s_i} = 0 + \beta\hat{\theta}_{s_i} + \varepsilon_i$ (model performance measures follow Janssen & Heuberger 1995).

Site	RMB	IOA	NMAE	Regression estimates		
				β	β stderr	r^2_{adj}
Mumbalup	-0.224	0.97	0.264	1.057	0.097	0.90
Manjimup	0.013	0.99	0.164	0.961	0.050	0.97

4.3.1.2 Eucalyptus globulus plantation sites in Northern Tasmania

For compiled ecological data, the available substrate information may only be an approximation to soil texture from parent rock types. Soil texture information is adequate for estimating a coefficient of evapotranspiration from the relevant soil water retention function, but does not indicate other site differences in maximum soil water storage due to variation in soil depth and rockiness. To determine the effects of limited information in the derivation of a gradient in soil water supply, several simulations of water balance with varying levels of substrate information were compared with the estimates of soil water relations derived from the genetic parameters of response for *E. globulus* plantations in Northern Tasmania (after Battaglia & Sands 1997).

Table 4.6 indicates the range in site conditions predicted for the 19 *Eucalyptus globulus* plantations in Tasmania, based upon the water balance model with substrate variation in texture, stone fraction and soil depth. These plantations were located in a climatic region with relatively high annual rainfall (Table 4.3), and distinct seasonality — higher winter rainfall and lower summer rainfall. The seasonality and monthly range of estimates for actual evapotranspiration, soil water potential and the index of surplus water are shown in Figure 4.3. Estimates of actual evapotranspiration from this model of water balance do not distinguish sources of evaporation. In forest vegetation, the majority of evaporation passes through plant stomata, in which case it is transpired-water, but periods of high vapour pressure deficits may promote stomatal closure (e.g. Breda *et al.* 1993; Tiktak & Bouten 1994), while evaporation continues from bare soil and other moist surfaces. Transpiration over annual periods is proportional to gas exchange for photosynthesis and, therefore, plant growth (Ball-Berry 1982; Zhang & Nobel 1996). As a result, actual evapotranspiration in forest vegetation is indicative of plant growth potential, albeit overestimated.

Table 4.6 Range in annual site conditions (mean, maximum, minimum) predicted across 19 *Eucalyptus globulus* plantations in Northern Tasmania using observed levels of substrate variation (after Laffan 1993; Osler *et al.* 1996). E_a is the actual evapotranspiration, E_a/E_o is the ratio of actual to potential evapotranspiration, Ψ_s is the soil water potential (MPa), $R^*_{\theta s}$ is the ratio of available soil water content to maximum available soil water storage and $P^*_{\theta R}$ is the ratio of water runoff to rainfall (after Church *et al.* 1995).

Site	E_a (mm/yr)	E_a/E_o (%)	Ψ_s (MPa)	$R^*_{\theta s}$ (%)	$P^*_{\theta R}$ (%)
Annual total	708	78	-0.3400	56	20
Monthly maximum	115	99	-0.0075	99	73
Monthly minimum	29	27	-1.1400	3	0

The seasonal variation in actual evapotranspiration for the *E. globulus* plantation sites (Fig. 4.3a) suggest that growth peaks during late spring and early summer, before the onset of drought in January. Early summer is also the period of highest solar radiation (peaking in late December). The seasonal variability in soil water potential and water surplus have inverse relationships (Fig. 4.3a, b). Site differentiation in water relations is apparent during the initial drying phases in November and December, where variation in moisture availability is greatest. In these months, substrate factors have more influence on the buffering of sites from the early onset of drought; but as summer progresses, climate has more influence over site water relations. Since actual evapotranspiration rates also peak during the same period (October to December, depending on the site), it is the buffering capacity of the substrate, rather than climate, that largely distinguishes the growth potential of these sites.

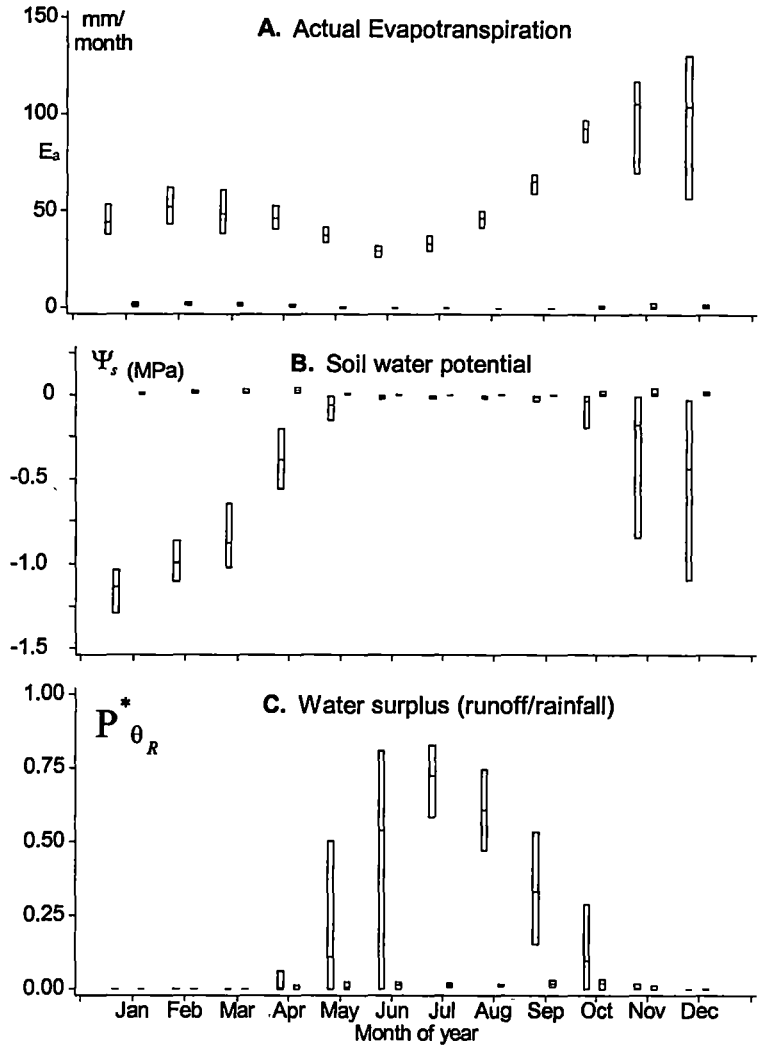


Figure 4.3 Seasonality and monthly range of actual evapotranspiration (A), soil water potential (B) and an index of surplus water (C) for 19 *Eucalyptus globulus* plantation sites in Northern Tasmania (see summaries of environmental data in Table 4.3).

Estimates for evapotranspiration, soil water potential and surplus water were calculated on a daily step and summed for the month, averaged over 10 annual cycles of rainfall per day, presented here as the mean and standard error (small adjacent bars). Bars indicate maximum, minimum and mean values.

The degree to which growth of *E. globulus* is limited by water stress on 19 plantation sites in northern Tasmania, estimated from the process model PROMOD (Battaglia & Sands 1997), is compared in Figure 4.4 with the mean annual soil water potential derived from the modelling of water balance using the soil water retention function. In this case growth of *E. globulus* from PROMOD is based on the more specific estimate of actual transpiration and is compared with a combined estimate derived through actual evapotranspiration, which does not distinguish evaporation from different sources. There is a reasonable degree of consistency between the two estimates of site water relations ($r^2_{adj} = 0.93$), indicating that substrate parameters (soil depth, soil texture and rockiness), can be used to approximate the ranking of site differences for the purpose of water balance modelling, where vegetation parameters are incompletely documented. This is because stomatal conductance and the soil water retention curve have similar functional shapes. Even though one function is a biological response, and the other an environmental response, species' responses are highly integrated with the soil environment. Therefore, for the purpose of defining a gradient in water supply that can be applied to a comparison of different species, the soil water retention function for estimating evapotranspiration directly correlates with the physiological responses of transpiration, photosynthesis and growth potential.

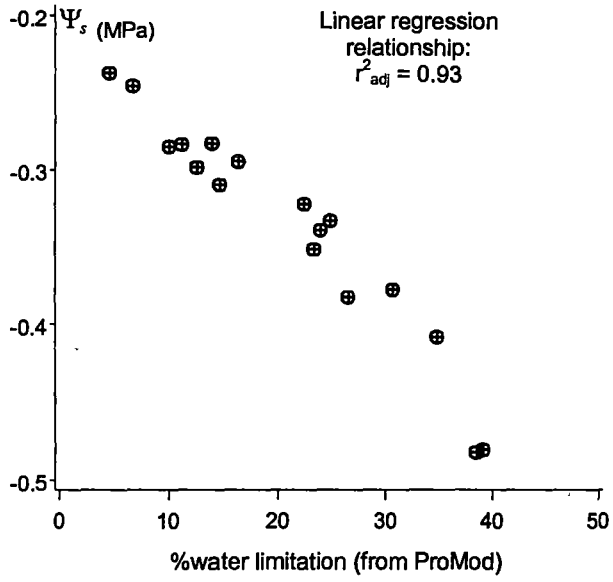


Figure 4.4 Comparison of the degree to which growth of *E. globulus* at a site is limited by water stress as estimated by PROMOD (Battaglia & Sands 1997), with soil water potential (Ψ_s) as estimated by the model for potential environmental water balance.

Both estimates of water stress are based on the same set of substrate characteristics described in Table 4.3.

In the absence of any substrate interactions, the effects of environmental water, approximated as the annual net climatic water balance (sum of the monthly differences between precipitation and evaporation), explains 73% (r^2_{adj}) of the variation in PROMOD's water-stress index. With substrate interactions, the modelled site water relations explained 93% of the PROMOD index. However, when substrate variation due to stone fraction is removed, the comparison reduces to 85% of the PROMOD index, and with all substrate conditions unvarying, this comparison is 81% (Table 4.7). These results indicate that the modelling of water balance, even where the substrate descriptions are incomplete, provides estimates for the ranking of site water relations that are better than an index of environmental water (precipitation minus evaporation), which does not allow for substrate interaction through a buffering of soil water storage. This is because substrate factors may be seasonally much more important than climatic factors (which dominate the average annual site water relations) (e.g. Fig. 4.3A).

Table 4.7 Comparison of estimates of water supply based on the soil water retention function with estimates of the degree to which growth of *Eucalyptus globulus* is limited by water stress (from the process model PROMOD; Battaglia & Sands 1997) for 19 plantation sites in northern Tasmania (after Osler *et al.* 1996). The adjusted r^2 (%) is derived from the linear regression relationship $y_i = 0 + \beta x_i + \varepsilon$ between % water limitation (from PROMOD) and each annual water balance variable (actual evapotranspiration, E_a; potential evapotranspiration, E_o, ratio of evapotranspiration, E_a/E_o; soil water potential, Ψ_s; soil relative water content, $R^*_{\theta S}$; index of soil water surplus, $P^*_{\theta_R}$; and soil water runoff, θ_r). The Pearson correlation coefficients (Corr, %) between each annual water balance index and % water limitation are also given. Comparisons are described in the methods.

Comparisons	Ψ_s		E_a/E_o		$R^*_{\theta S}$		$P^*_{\theta_R}$		E_a		θ_r	
	r^2_{adj}	Corr	r^2_{adj}	Corr	r^2_{adj}	Corr	r^2_{adj}	Corr	r^2_{adj}	Corr	r^2_{adj}	Corr
1. Substrate variation	93	–97	91	–96	81	–91	15	–47	72	–87	39	–66
2. Soil depth = 100 cm	94	–97	94	–97	72	–86	33	–61	82	–91	48	–71
3. No stones, 100 cm	84	–92	84	–92	69	–84	53	–75	56	–76	57	–77
4. Texture = 'clay'	84	–92	85	–93	81	–91	51	–74	58	–78	57	–77
5. Substrate constant	81	–91	81	–91	77	–88	72	–86	49	–72	66	–83

While these comparisons are dataset-specific, they are typical of forest stands (M. Battaglia personal communication) and are indicative of the probable sensitivities to data quality in the modelling of water balance. The two different approaches to modelling water balance (soil environment versus plant physiological response) were also reasonably comparable because site differences in maximum soil water storage, substrate depth and rockiness have a major influence on water relations in northern Tasmania (over the physiological response of *E. globulus* to variation in water supply between sites). Substrate factors are especially important during the growing season of late spring to early summer. That is, the brief season between the cold/short-day conditions which limit growth in winter and the drought limitations to growth in summer. The importance of substrate compared with climate declines during either the core winter or summer periods, when the soils are either constantly saturated or predictably prone to drought. However, limitations in the level of substrate information available in compiled ecological data only reduced the precision of estimating site water relations by about 10% (from 93 to 84% between comparisons 1 and 3 in Table 4.7).

The water balance model splits environmental water into two components, soil water potential and soil water runoff. Soil water potential represents a water supply gradient which is available for evapotranspiration and is strongly correlated with vegetation processes. Soil water runoff (combination of drainage by deep percolation and surface runoff), however, is generally surplus to the capacity for storage at the site, and is (by definition) lost to the plant environment. This split of environmental water was maximised when all observed substrate variation was considered in the water balance model for the 19 *E. globulus* plantation sites (Table 4.8). For example, the Pearson correlation coefficients between two indicative variables, mean annual soil water potential and total annual soil water runoff were about 51% (see comparison 1). However, where the level of site information which was generally available from compiled ecological data

was considered (see comparison 3), the correlations increased to about 78%. Where no site variation was assumed, the correlations were greatest (96%, see comparison 5).

Table 4.8 Pearson correlation coefficients (with Prob > |R| under Ho: Rho=0) between indicative variables from the water balance model (soil water potential, Ψ_s; and soil water runoff, θ_r, index of soil water surplus, $P^*_{\theta_R}$) for the five comparisons and 19 sites in Table 4.7.

	Ψ_s				
Comparisons	1	2	3	4	5
θ_r	0.50662	0.67410	0.78309	0.74768	0.95711
Prob > |R|	0.0269	0.0016	0.0001	0.0002	0.0001
$P^*_{\theta_R}$	0.23852	0.52394	0.71812	0.64812	0.97451
Prob > |R|	0.3254	0.0213	0.0005	0.0027	0.0001

These correlations suggest that the water balance model was inaccurately partitioning environmental water when the substrate information was restricted to a common set of conditions between sites (e.g. one metre soil depth, no rocks). Hence two components of the water supply gradient (e.g. soil water potential and soil water surplus) may need to be considered when developing models of species' performance with compiled ecological data. For example, the variable for soil water runoff may reflect information which is better at discriminating site differences in water relations for some species, than either the soil water potential or the index of evapotranspiration.

4.3.1.3 Water balance variation in Tasmanian eucalypt forests

The substrate information available for modelling water balance with the set of data for eucalypt distributions in Tasmania comprised only an approximation of soil texture type from parent rock categories. These approximations were consistent with the perceived weathering potential of parent rock types and the experience of forest soil scientists in different climatic regions (e.g. Grant *et al.* 1995). However, because the soil environment varies (e.g. more or less than one metre of stone-free depth, more than one soil layer, intermediate or different substrate textures to sand, clay and loam types), the water balance model inaccurately partitioned environmental water between soil storage and surplus. As a result, the model of water balance largely distinguishes site variation due to climate. The water relations of a site will only be partially defined by soil water potential and additional information from the variable for soil water runoff may be needed to appropriately rank water balance differences between sites.

The seasonality and monthly range of soil water potential (MPa) and soil water runoff (mm) modelled for the range of forest habitats in the ecological dataset is shown in Figure 4.5. Most sites have saturated soil profiles over the winter months which successively dry out over the spring and summer months. Relatively few sites are not saturated over winter, and few do not markedly dry out over summer. There is a broad range in variation between sites during the summer months and a narrower range in winter. This seasonal variability in water relations

associated with each site may be summarised as statistics for the mean annual, maximum and minimum months. These statistics represent different indices for water supply.

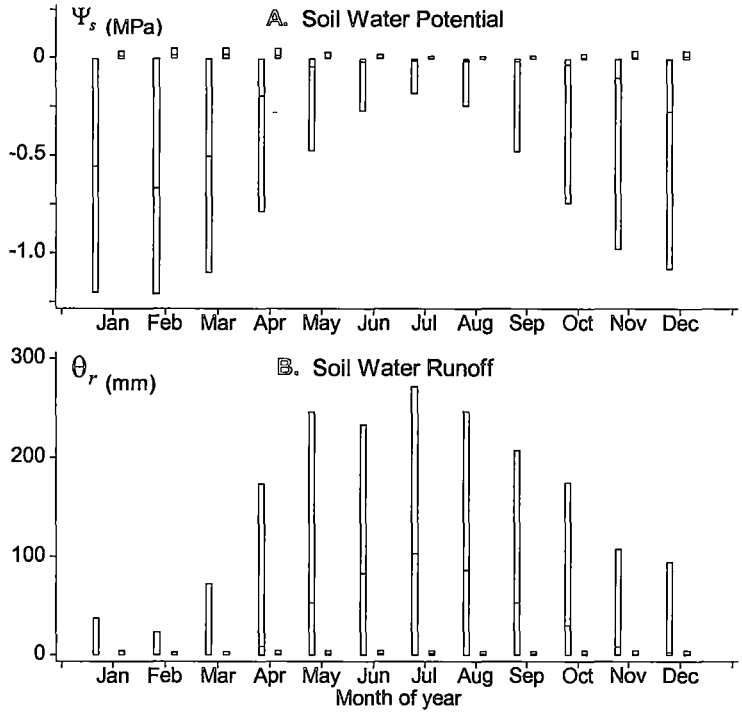


Figure 4.5 Seasonality and monthly range of soil water potential (𝔸) and soil water runoff (𝔹) for a range of forest habitats dominated by *Eucalyptus* species in Tasmania (n = 15640).

Estimates are calculated on a daily basis for one metre of soil depth (varying by three classes of soil texture and five classes of available water), summed for the month, averaged over 10 annual cycles of rainfall per day, and presented here as the mean and standard error (small adjacent bars). Bars indicate maximum, minimum and mean values.

The envelope of site variation in mean annual soil water potential and total annual soil water runoff is presented in Figure 4.6, showing the overlap and variation associated with the three soil texture classes. The Pearson correlation coefficients indicate a reasonable dependence between these two variables (about 79% correlated). This was consistent with the 78% correlation between these variables from the modelling of 19 *E. globulus* sites when substrate variation was also reduced to texture differences only (soil depth was constant at 100 cm and stone fraction was negligible for comparison 3 in Tables 4.7 & 4.8).

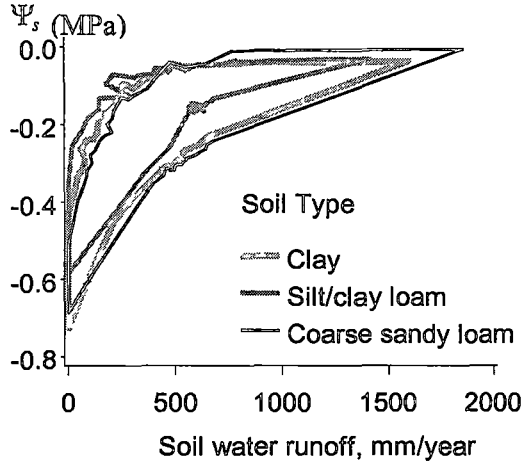


Figure 4.6 The relationship between two indices for water balance, mean annual soil water potential and total annual soil water runoff by the three texture classes water for a range of forest habitats dominated by *Eucalyptus* species in Tasmania (n = 15 640).

Pearson correlation coefficients (%) between soil water potential and runoff for clay, silt/clay loam, coarse sandy loam and the complete data range are 73.1, 85.0, 84.2 and 78.5% respectively.

The table of correlations between the mean annual statistics for three environmental water variables (rainfall, evaporation and their net difference) and two water balance variables (soil water potential and soil water runoff) demonstrates the altered relationship between water and temperature following the modelling of water balance (Table 4.9). For example, the relatively close relationship between evaporation and temperature (72 to 78%), is more distant when water is redefined with respect to substrate as soil water potential (52 to 58%) and runoff (37 to 44%). The relatively uncorrelated relationship between rainfall and temperature (26 to 35%) is partially carried through the modelling of water balance to the variable for soil water runoff. Soil water potential is approximately equally related to either evaporation or rainfall (78%), but soil water runoff is strongly correlated with rainfall (97%). Net climatic water is also dominated by the rainfall component (95%).

Table 4.9 Correlations between water balance estimates and climatic variables for a range of forest habitats dominated by *Eucalyptus* species in Tasmania (n = 15 640). Pearson correlation coefficients between environmental water variables (total annual mean precipitation, P_T; total annual mean pan evaporation, E_P; total annual net climatic water, W_{P-E}), water balance variables (mean annual soil water potential, Ψ_s; total annual soil water runoff, θ_r) and temperature (mean annual mean daily minimum and maximum temperatures, T_i & T_x).

	θ_r	Ψ_s	P_T	E_P	W_{P-E}	T_i	T_x
θ_r	1.00000	0.78487	0.97171	–0.46511	0.95344	–0.36953	–0.43624
Ψ_s		1.00000	0.77543	–0.78216	0.89710	–0.51667	–0.58469
P_T			1.00000	–0.39163	0.95234	–0.26181	–0.34505
E_P				1.00000	–0.65363	0.71704	0.77610
W_{P-E}					1.00000	–0.45304	–0.54109
T_i						1.00000	0.89185
T_x							1.00000

Soil water potential was expected to be the main component of environmental water and soil water runoff was generally expected to contribute a residual component. However, either component may be more important for a species' response. Which is more important depends on the balance between rainfall and evaporation buffered through one metre depth of soil at a site, and the actual soil environment. The correlations indicate that either component of water balance (soil water potential or runoff) or both may be significant when modelling a species' response which will also depend upon the relative importance of variation due to temperature, and other independent variables for light and nutrients.

4.3.2 Modelling performance of individual *Eucalyptus* species

Logistic regression models were used to compare the importance of a water supply gradient, derived from the modelling of water balance with limited substrate information, relative to climate factors for the distribution of four *Eucalyptus* species in Tasmania (a general description of the habitat associated with each species' distribution is given in Williams & Potts 1996). The change in deviance in a logistic regression for the univariate fit of these different indices of annual or seasonal water supply (summarised as statistics for the mean annual, maximum and

minimum months) is summarised in Table 4.10. The different response of each species was reflected by the variables contributing the larger change in deviance from the null model (significant at the 0.01% level). The expectation from the modelling of water balance, even with limited site information, was that for any species a single gradient in water supply would result in a greater change in deviance than any single climatic factor for water. This expectation also arises from ecological theory which states that environmental gradients appropriate to the study of the continuum concept should be directly related to plant physiological responses (Austin & Smith 1989). A gradient in soil water supply was considered to be more closely related to plant physiological processes than a gradient in rainfall or evaporation.

Table 4.10 Univariate ecological responses to water. Change in deviance (logistic regression) for variables defining climatic water conditions or modelled water relations for observations of the presence (n_1) and absence (n_0) of four *Eucalyptus* species. The degrees of freedom (df) indicate the number of polynomial elements (up to the fourth order) included in the univariate model. Results are only shown for variable fits which are significant at the 0.01% ($p < 0.0001$) level or greater (else fits are not significant: *ns*). The most significant responses for each species in each variable set is indicated in bold.

	E. regnans ($n_1 = 2462$) ($n_0 = 6713$)		*E. obliqua* ($n_1 = 8182$) ($n_0 = 5433$)		*E. amygdalina* ($n_1 = 3986$) ($n_0 = 8321$)		*E. globulus* ($n_1 = 1091$) ($n_0 = 3980$)	
Variable	ΔDev (NULL = 10 672)	df	ΔDev (NULL = 18 316)	df	ΔDev (NULL = 15 501)	df	ΔDev (NULL = 5280)	df
Climatic water variables:								
Pan evaporation (E_P, mm/month)								
Total annual	**617**	4	495	4	3125	3	92	3
Maximum month	524	4	647	4	**3474**	4	53	2
Minimum month	479	4	**672**	4	2210	3	**120**	3
Rainfall (P_T, mm/month)								
Total annual	498	2	731	3	1327	4	215	4
Maximum month	195	4	**901**	3	344	3	**236**	4
Minimum month	**714**	3	313	4	**2242**	4	153	4
Raindays (R_D, days/month)								
Total annual	361	4	719	4	2747	4	ns	ns
Maximum month	**422**	2	881	4	1642	4	50	3
Minimum month	391	4	**983**	4	**3437**	4	43	4
Net climatic water balance (Rainfall-Pan evaporation) (W_{P-E}, mm/month)								
Total annual	**747**	3	682	4	2372	4	167	4
Maximum month	272	4	**1034**	3	533	3	**192**	4
Minimum month	579	2	382	4	**3510**	3	81	2
Processed Water-Balance variables:								
Soil Water Potential (Ψ_s, MPa)								
Mean annual	**827**	2	**584**	4	3018	3	**111**	2
Minimum month	726	4	139	3	**3178**	3	101	4
Soil Water runoff (θ_r, mm/month)								
Total annual	**449**	2	671	4	**1381**	4	**199**	4
Maximum month	263	2	**1051**	2	523	3	156	4

The univariate responses of the four *Eucalyptus* species to water balance or climatic factors are varied (Table 4.10). For two species, *E. obliqua* and *E. regnans*, a water balance variable was more important than a climatic variable in describing their response. This was not the case for *E. globulus* or *E. amygdalina.*

Only *E. regnans* demonstrated a strong relationship with the anticipated water supply gradient, mean annual soil water potential, and the shape of this response was best defined by a quadratic function, representing a classic symmetric response. This response may be because the presumed substrate conditions (one metre soil depth with texture variation) were reasonable assumptions throughout the range of this species. However, this possible coincidence between actual and presumed substrate conditions did not occur for the remaining species.

While the water balance variable for soil water runoff represents a slightly better fit to the response of *E. obliqua*, this is not much different to a variable for net climatic water balance (Δ Deviance: 1051 versus 1034). Net climatic water balance, calculated without any substrate interactions, also represents the best fit for the response of *E. amygdalina*, but for a contrasting type of habitat compared with *E. obliqua* (high evaporation and low rainfall versus low evaporation and high rainfall). The distribution of *E. obliqua* is likely to be associated with soils of varying depth, depending upon the interaction with climate and competition with *E. regnans* or rainforest species on the best sites (deep soils, warm, moist habitats). *Eucalyptus amygdalina*, however, is likely to be associated with shallow soils wherever it occurs, or deeper sandy soils with very low water holding capacities. This contrasts with *E. globulus* whose response is mainly influenced by rainfall rather than water balance estimate (Δ Deviance: 236 versus 199).

Of all the species investigated here, *E. globulus* may be the most sensitive to local substrate conditions, relying on the interaction between climate and soil water storage to buffer its occurrence in habitats which might otherwise appear too dry (e.g. White *et al.* 1996). Its widespread distribution and low abundance in the dry sclerophyll forests of south-eastern Tasmania may also reflect a tendency to occupy such microhabitat situations, surrounded by the more abundant series *Piperitae* species, such as *E. amygdalina* (Duncan & Brown 1985).

These univariate responses clearly demonstrate the importance of the balance between evaporation and rainfall in determining species' distributions. The comparisons, however, also highlight the inadequacy of generating gradients in soil water supply based on incomplete substrate information, which confounds the ranking of sites by their average water relations. The previous analyses had indicated that substantial losses in information about the appropriate ranking of sites along a water supply gradient can be expected with accumulated assumptions of common soil conditions.

The potential response of each species to the combined effects of temperature and water is indicated in Table 4.11, and the corresponding model fit statistics are given in Table 4.12. These models test whether gradients in water balance were significant improvements over gradients in

climatic water, when the correlations with temperature were also taken into account. The results confirm the earlier observation (Table 4.10) that the variables derived from the water balance model may not describe site variation as comprehensively as the original climatic variables for environmental water. For each species, a model derived from climatic factors (rainfall, rain days, evaporation and temperature) is a better fit to its response (see Δ deviance) than a model derived from the two components of water balance (soil water potential and soil water runoff). However, models derived from climate estimates were generally more complex, involving a larger number of variables.

Table 4.11 Multivariate ecological responses to water and temperature. Potential maximum response of each species to total annual variables of climatic water (rainfall, P_T; evaporation, E_P, and rain days, R_D) compared with the modelled water balance variables (mean annual soil water potential, Ψ_s; total annual soil water runoff, θ_r) in the presence of a temperature variable (mean annual minimum or maximum temperature, T_i & T_x). The linear predictor (η) is given for the fit of the logistic regression functions. The NULL model statistics and number of observation (presences and absences) are given in Table 4.10. Model fits are indicated in Table 4.12.

Model	
Eucalyptus regnans:	
Climatic water conditions, η =	$-619.1 + 1.72\times10^{-2}\times P_T - 5.96\times10^{-6}\times P_T^2 + 2.7884\times E_P - 4.97\times10^{-3}\times E_P^2 + 3.873\times10^{-6}\times E_P^3 - 1.12\times10^{-9}\times E_P^4 + 0.2267\times R_D - 7.5\times10^{-4}\times R_D^2 + 4.9084\times T_i - 0.3895\times T_i^2$
Modelled water relations, η =	$-14.7407 - 8.8064\times\Psi_s - 34.8946\times\Psi_s^2 + 3.88\times10^{-3}\times\theta_r - 4.1\times10^{-6}\times\theta_r^2 + 4.7900\times T_i - 0.4288\times T_i^2$
Eucalyptus obliqua:	
Climatic water conditions, η =	$-755.5 + 0.0339\times P_T - 0.03\times10^{-3}\times P_T^2 + 6.792\times10^{-9}\times P_T^3 - 0.0423\times E_P + 2.3\times10^{-5}\times E_P^2 + 11.5602\times R_D - 0.0940\times R_D^2 + 3.39\times10^{-4}\times R_D^3 - 4.57\times10^{-7}\times R_D^4 + 41.1401\times T_x - 2.4701\times T_x^2 + 0.0496\times T_x^3$
Modelled water relations, η =	$-64.8541 + 35.5153\times\Psi_s + 198.4\times\Psi_s^2 + 413.6\times\Psi_s^3 + 272.1\times\Psi_s^4 + 9.05\times10^{-3}\times\theta_r - 0.01\times10^{-3}\times\theta_r^2 + 5.515\times10^{-9}\times\theta_r^3 + 8.2445\times T_x - 0.2562\times T_x^2$
Eucalyptus amygdalina:	
Climatic water conditions, η =	$-1343.1 + 0.0302\times P_T - 0.03\times10^{-3}\times P_T^2 + 7.399\times10^{-9}\times P_T^3 + 0.0145\times E_P + 21.8330\times R_D - 0.1807\times R_D^2 + 6.56\times10^{-4}\times R_D^3 - 8.82\times10^{-7}\times R_D^4 + 72.4317\times T_x - 5.0468\times T_x^2 + 0.1156\times T_x^3$
Modelled water relations, η =	$+433.2 - 27.6195\times\Psi_s - 93.2991\times\Psi_s^2 - 240.1\times\Psi_s^3 - 215.7\times\Psi_s^4 + 0.0229\times\theta_r - 0.06\times10^{-3}\times\theta_r^2 + 6.492\times10^{-8}\times\theta_r^3 - 2.15\times10^{-11}\times\theta_r^4 - 145.5\times T_x + 17.3862\times T_x^2 - 0.8987\times T_x^3 + 0.0170\times T_x^4$
Eucalyptus globulus:	
Climatic water conditions, η =	$-267.0 + 0.3639\times P_T - 5.5\times10^{-4}\times P_T^2 + 3.574\times10^{-7}\times P_T^3 - 8.55\times10^{-11}\times P_T^4 - 0.0668\times E_P + 0.31\times10^{-4}\times E_P^2 + 124.8\times T_i - 27.0513\times T_i^2 + 2.5820\times T_i^3 - 0.0914\times T_i^4$
Modelled water relations, η =	$-183.7 + 7.8349\times\Psi_s - 5.72\times10^{-3}\times\theta_r - 0.05\times10^{-3}\times\theta_r^2 + 1.684\times10^{-7}\times\theta_r^3 - 1.4\times10^{-10}\times\theta_r^4 + 112.5\times T_i - 25.0113\times T_i^2 + 2.4267\times T_i^3 - 0.0864\times T_i^4$

Table 4.12 Multivariate ecological responses to water and temperature. Comparison of fits to statistical models for climatic or modelled water relations. The NULL model statistics and number of observation (presences and absences) are given in Table 4.10. The model equations are given in Table 4.11. Response shapes for each variable are defined by the order of the polynomial. Results are only shown for fits which are significant at the 0.01% ($p < 0.001$) level or greater, following backward selection of the maximum number of variables. The model fits are indicated by the *Degrees of Freedom* (df), *Change in Deviance* (ΔDev), the *Association of Predicted Probabilities and Observed Responses* (concordant, discordant and tied responses), and the *Classification Table for Trials of Event/Noevent* (based upon the response being an event if the probability of occurrence is greater than 0.5, indicating the percentage correct, and rates of false positive, F+, and false negative, F–, predictions). Details of the logistic regression method are given in SAS Institute Inc. (1990d).

Species	df	ΔDev	conc	disc	tied	%correct	F+	F–
Eucalyptus regnans:								
Climatic water relations	10	1591.6	76.1	23.6	0.3	75.2	41.6	22.4
Modelled water relations	6	1365.2	73.7	26.0	0.3	73.8	43.7	25.3
Eucalyptus obliqua:								
Climatic water conditions	12	4171.3	80.1	19.6	0.3	74.2	26.2	24.9
Modelled water relations	9	3363.5	76.4	23.3	0.3	72.6	27.6	27.0
Eucalyptus amygdalina:								
Climatic water conditions	11	4355.1	83.9	16.0	0.2	78.0	30.9	18.6
Modelled water relations	12	4084.9	82.7	17.1	0.2	78.3	28.6	19.5
Eucalyptus globulus:								
Climatic water conditions	10	492.7	70.3	29.2	0.5	78.5	36.4	21.4
Modelled water relations	9	420.4	66.9	32.5	0.6	78.4	52.2	21.2

4.4 Discussion

A simple model of site water balance, based on the soil water retention function, was developed to define a gradient in soil water supply that is more directly related to the physiological response of plant species than gradients in climate or categories of substrate. The elaboration of this model was limited by the level of information available with compiled ecological data. The choice of an appropriate method for estimating water balance therefore assumed that a water supply gradient need only distinguish sites in the appropriate order of their annual and seasonal differences in water relations. Accurate estimates for the parameters of water balance were not considered a priority. Therefore, substrate differences such as soil texture, structure, depth and rockiness and their interaction with climate were assumed to be of primary importance for distinguishing relative site differences in water relations. Vegetation attributes such as leaf area index and canopy conductance were not known, and assumptions for these parameters were not considered appropriate for deriving a water supply gradient to compare the response of different *Eucalyptus* species.

Comparison between observed and simulated estimates of soil water content demonstrated the utility of a simple model of soil water balance based on the soil water retention function. This relationship between soil water potential and soil water content is a characteristic property of soils of different texture which can be easily approximated from a text book (e.g. Taylor & Ashcroft 1972). However, it increasingly over-predicted the daily drying rate of *E. globulus* plantations as the mean annual rainfall declined. This deviation may be due to a non-linear

functional relationship between the vegetation response and the soil environment, since a proportional relationship was assumed in developing this model of water balance. For example, these differences could be related to the restriction of stomatal conductance when root-zone matrix potentials fall below a threshold value (e.g. Whitehead & Kelliher 1991), and deep water reservoirs could also contribute to a buffering effect for the plantation during periods of low rainfall (Hingston *et al.* 1994). For simplicity in application, the modelled soil water contents were approximated for a single layer root-zone of clay texture, rather than the multilayered soil structure which existed at each site (Hingston *et al.* 1994). A closer matching between observed and predicted daily soil moistures would be likely for experimentally-determined soil water retention curves modelled for each substrate layer (e.g. Walker & Langridge 1996). Despite these differences, the soil water retention function realistically simulated the general process of water relations, with much the same level of accuracy as a model based on the physiological response of *E. globulus* (Battaglia & Sands 1997).

The results of modelling water balance for the 19 *E. globulus* plantation sites in Northern Tasmania showed that climatic factors dominated the average annual site water relations associated with these data, but that substrate factors (soil depth, texture and rockiness) were seasonally much more important in determining periods of growth. As the geographic scale at which site differences in water relations are compared is extended across major climatic regions, then the importance of climate over substrate factors is also likely to increase (e.g. Woodward & Williams 1987; Tchebakova *et al.* 1993; Huntley *et al.* 1995; Siegel *et al.* 1995a, b). Therefore, the response of widespread *Eucalyptus* species, such as *E. obliqua* or *E. amygdalina* could be more readily explained by mean annual climate factors than other species, such as *E. globulus* or *E. regnans*, which are regionally or locally restricted in their range. However, substrate effects on water relations may be increasingly important to a correct ranking of sites for distinguishing the response of species that extend into drought-prone regions (e.g. Kirkpatrick & Marks 1985; Davidson & Reid 1989).

The value of estimates of site water relations for predicting species' distributions was assessed by comparing the response of *Eucalyptus* species from contrasting habitats in logistic regression models (Tables 4.10 & 4.11). Since previous analyses had validated this method of modelling water balance, the reduced effectiveness of a water supply gradient was attributed to limited information about the substrate. As a result, some sites were incorrectly ranked relative to others. These errors in estimating water relations would have largely arisen from the assumption of constant soil depth. Across the range of sites in the compiled ecological data, the depth of soil profile makes a far greater contribution to the calculation of available water capacity than does the soil texture. Other errors might include the lack of consideration given to terrain influences on the redistribution of water across the landscape. One consequence of the soil depth assumption is the inaccurate allocation of environmental water between storage in the soil and

runoff from saturated surfaces. Some of this information was recovered by considering the variable for soil water runoff as an additional gradient with respect to water supply when building ecological models of a species' performance (Table 4.11). However, multivariate responses derived from climatic factors always gave better fits than the water balance variables, but these were complicated and could be more difficult to interpret. The relative importance of substrate and climate in water relations is therefore site-dependent, although atmospheric differences in water availability are likely to dominate annual variation in water relations between sites (e.g. Centurion *et al.* 1992). Where the results of ecological models are intended for comparison with experimental work, then the consideration of a species' response to a gradient in soil water supply, although a slightly poorer fit, may facilitate the biological interpretation of performance. The requirements for either or both water balance variables will, of course, depend upon the presence of other environmental factors (related to light, temperature and nutrients) in a model of relative plant performance.

The relative improvement in defining a water supply gradient will depend largely on the resolution of the substrate information. With a reasonable level of substrate information (texture, structure, depth and stone fraction), the water characteristics of the soil provided a mechanism for defining a water supply index which approaches a resource gradient (*sensu* Austin & Smith 1989). However, in the absence of more complete information for site differences in soil texture, structure and volume, or genetic parameters for the plant physiological response, a soil moisture gradient of the type defined for the compiled set of ecological data must be considered an indicator, only, of resource supply, but is more like a direct environmental gradient (Austin 1990). Future developments in remote sensing may enable the appropriate parameters for substrate moisture relations to be indirectly measured (e.g. Ragab 1995). Remote sensing, combined with physical models of soil properties (e.g. Petersen *et al.* 1996; Kolev *et al.* 1996) and maps of soil texture type (e.g. Barringer *et al.* 1995) or their prediction from terrain (e.g. Merot *et al.* 1995; Vertessy *et al.* 1993) could further contribute to more accurate estimation of site water relations.

An alternative approach might be to utilise published sources of soil attribute information from land surveys. This idea was explored by transcribing the basic terran and soil attribute information available with the published survey reports of the land systems in Tasmania (e.g. Pemberton 1986). However, the subsequent estimates of site water balance resulted in an environmental gradient that was a poorer predictor of species' distribution patterns, than original estimates that assumed constant soil conditions. It was therefore decided that broadly-mapped estimates of site substrate conditions would not be helpful for species distribution modelling. The broad differences in scale and the generalised descriptions in the text of each land system report probably contributed to the inaccuracy of these water balance estimates. However, it may be possible to improve such site estimation by developing a model of landscape soil processes from the original site information. Other surveys of Tasmanian forest soils that have been mapped at

finer scales (e.g. 1:100 000 maps by Grant *et al.* 1995a; Laffan *et al.* 1995; Hill *et al.* 1995), could be applied to an approximation of substrate characteristics in compiled ecological data. The potential for these more detailed surveys and mapping studies to increase the precision of modelling water balance for compiled ecological data needs to be tested. For example, the increasing resolution of geological mapping for Tasmania (e.g. 1:250 000 series with over 300 categories), combined with the descriptions of forest soils on different parent rock types (e.g. Grant *et al.* 1995b) could further increase the accuracy of approximations of soil texture class, and improve the resolution of water balance estimates.

4.5 Conclusions

A model of site water balance, based upon the potential limitations imposed by the environment, and without specific reference to vegetation parameters, was found to be a reasonable approach to developing a direct gradient in water supply for ecological analyses. A validation study of the soil water retention function as a method for defining the relationship between evapotranspiration and soil water content was conducted across several scales and investigated different levels of substrate information. The results demonstrated that the effort involved in modelling water balance should reflect the accuracy or resolution of the substrate information in compiled ecological data, particularly soil depth and texture characteristics which ultimately determine the water relations of a site. Such information may be available from specific studies, but was not consistent between sources which contributed to the compilation of ecological data for *Eucalyptus* species' distribution in Tasmania. The application of 'best-guess' estimates of soil type for broad classes of geological type did not improve the derivation of ecological indices for the modelling of species' performance. However, integrated soil maps at a reasonable resolution, or models of soil genesis from landscape attributes, could improve the ranking of sites.

Models of species' distributions that are based on incomplete estimates of site water relations may result in poorer predictions than models based on the original climate data. However, corroborating evidence from experimental studies adds credibility to the ecological interpretation of plant distributions. An example of the application of a water supply gradient for comparison between observational and experimental data for the response of co-occurring *Eucalyptus* species is given in the next chapter.

5. A case study: mixed-species stands of eucalypts as ecotones on a water supply gradient[1]

5.1 Introduction

This chapter demonstrates the use of a simple water balance model to construct a direct environmental gradient of water stress. A gradient in water supply allows species' distribution patterns from inventory data to be compared with experiments conducted at different environmental scales.

The continuum concept proposes that vegetation varies continuously with variation in the environment (Gauch & Whittaker 1972; Austin 1985; Austin & Smith 1989). However, the continuum concept refers to gradients in abstract environmental space, and consequently, homogeneous vegetation assemblages may be perceived where physical environmental gradients are shallow (Austin & Nicholls 1988). In these situations competitors may coexist as a result of chance dispersal or disturbance (e.g. Pickett & Bazzaz 1978; Turner *et al.* 1991), stochastic variation in regeneration conditions between recruitment events (e.g. Lamont & Witkowski 1995) or local-scale variation in an environment that is broadly homogeneous at a larger scale (e.g. Collins & Good 1987; Collins 1990).

Studies of the patterns of *Eucalyptus* species' association in Australia have found a significant reduction in the probability of codominant species coming from the same subgenus compared to what would be expected from random association (Austin *et al.* 1983). This observation has been termed "Pryor's rule" (Pryor 1953, 1959; *cf.* Austin *et al.* 1983; Noble 1989) and states that mixed-species stands of eucalypts usually will be composed of members from different subgenera. Nevertheless, broad areas over which members of the same subgenus co-occur are not uncommon. In some of these cases, resource partitioning between species has been demonstrated (Rogers & Westman 1979). In the remaining cases it has been suggested that shallow physical gradients are acting to diminish the influence of competitive hierarchies, and that these mixed-species stands are in fact broad ecotones (Gill & Ashton 1971; Ashton 1981; Florence 1981).

The role of drought-stress tolerance and avoidance in defining the distribution of parapatric *Eucalyptus* species, and leading to the formation of a landscape composed of a mosaic of repeated species' patches has been well demonstrated (e.g. Sinclair 1980; Clayton-Greene 1983; Davidson & Reid 1985). However, the interaction between differences in species' drought-stress tolerance and microscale variation (at the scale of metres to tens of metres) in the environment where species' co-occur has received less attention. In particular, few studies dealing with sets of co-occurring *Eucalyptus* species have attempted to provide a physiological explanation for the

[1] This chapter is based on material previously published as Battaglia and Williams (1996)

changes in plant performance along drought-stress gradients; nor have they linked this to environmental heterogeneity within the ecotonal zone (although see Sharitz & McCormick (1973), Platt (1975), Pickett & Bazzaz (1978), Shipley & Keddy (1987), Dawson (1990), Pantastico-Caldas & Venable (1993) and Burton & Bazzaz (1995) for examples of some parts of this with other plant genera).

This chapter considers whether the apparent mixed-species stands of two *Eucalyptus* species from the sub-genus *Monocalyptus* (Pryor & Johnson 1971), *E. obliqua* L'Her. and *E. tenuiramis* Miq., are broad ecotonal zones on a water-supply gradient. Whether environmental patchiness within this zone leads to dominance by one or other of the species (i.e. whether mixed-species stands are an artifact of the landscape examined (Austin & Smith 1989)) is tested. By examining the carbon economy of the two species along an artificial water-supply gradient, and by observing stand development in the field, the fundamental niche of the two species is explored and related to their ecological or realised niche. In doing so, the proposition of Austin and Smith (1989, p. 40) that "the species' fundamental response curves will be such that in a particular portion of the resource space a species will have a relative performance advantage over other species" is tested.

5.2 Methods

5.2.1 Database analysis of co-occurrence

Eucalyptus obliqua has an extensive distribution within Tasmania that totally encompasses the geographic range of *E. tenuiramis* (Williams & Potts 1996). *Eucalyptus obliqua* is widespread and common throughout Tasmania on moist (> 1000 mm rainfall per annum), well-drained lowland (< 600 m above sea-level, with infrequent occurrences to 800 m) sites (e.g. Duncan & brown 1985; Kirkpatrick *et al.* 1988a). It is a frequent dominant of wet forest, where it often occurs in pure stands, but it also extends into drier situations where it usually forms mixed-species stands with members of the subgenus *Symphyomyrtus*. *Eucalyptus tenuiramis* is endemic to south-eastern and eastern Tasmania where it occurs on insolated sites subject to moderate drought-stress. It is a frequent dominant of lowland dry forest below 600 m in altitude, with rare occurrences up to 700 m (Williams & Potts 1996). It usually occurs in mixed-species stands in association with members of the subgenus *Symphyomyrtus* but it tends to form pure stands on very dry sites.

In analysing the environmental domains of these two species, the study area was constrained to the geographic domain across which both species occur. That is, the analysis was restricted to plots (0.1 to 0.3 ha) that occur below 900 m in the south-eastern and eastern Tasmanian biogeographic regions 7a, 7b and 10b (*sensu* Orchard 1988). Within this study domain the database contained 2140 records in which either pure or mixed-species stands of *E. tenuiramis* and *E. obliqua* were defined. Pure stands were defined for this study as stands containing only

one of *E. tenuiramis* or *E. obliqua* with no other member of the subgenus *Monocalyptus* present. Mixed-species stands contained both *E. tenuiramis* and *E. obliqua*, but with no other member of the subgenus *Monocalyptus* present. The frequency of these data in each stand type is shown in Figure 5.1.

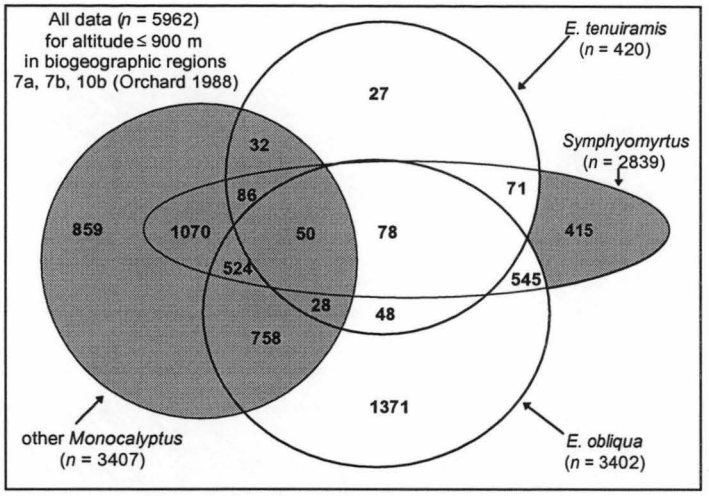


Figure 5.1 Set diagram indicating the proportion of the data in mixed-species and pure stands. The four larger ovoids indicate those plots containing *E. tenuiramis*, *E. obliqua*, other species in subgenus *Monocalyptus*, and species in subgenus *Symphyomyrtus*. The numbers enclosed by the *arcs* of the overlapping circles indicate the number of plots of each stand type in the dataset. Those stands used in the database analysis are unshaded.

Long-term monthly mean climatic data for precipitation, evaporation and temperature were derived using the process model ESOCLIM (developed by H. A. Nix, J. R. Busby, M. F. Hutchinson & J. McMahon; see description in Busby (1991)). These estimates of precipitation and evaporation were combined with soil characteristics to give estimates of soil water supply using a modified form of the simple water-balance model WATBAL (McAlpine 1970, Keig & McAlpine 1974). Soils were grouped into three broad texture classes (clay, silty clay-loam, coarse sandy-loam) and the water release characteristics of these respective classes used to convert soil relative water content to soil water potential (see Taylor & Ashcroft 1972, pg. 25) and to determine the field capacity and wilting point (Williams 1983).

The WATBAL model calculates the weekly change in water held in the soil as the balance between rainfall inputs and losses due to evapotranspiration and run-off. Rainfall in excess of the maximum soil water store, determined by soil texture and soil depth, is considered to be run-off. In our application of the model, the potential evaporation was estimated from slope and aspect adjusted open-pan evaporation derived from ESOCLIM. This was converted to potential evapotranspiration using a crop factor (PETCF in McAlpine 1970), and then to actual evapotranspiration (AETCF in McAlpine 1970) as the soil-water deficit increases and reduces canopy conductance. The program was stepped on a weekly time scale by dividing monthly estimates of rainfall and evaporation from ESOCLIM. The weekly outputs were averaged to give a monthly estimate of volumetric soil water content, and by application of a function for each soil texture class to give an estimate of soil water potential. All algorithms were the same as McAlpine (1970) except that AETCF was related to the proportion of remaining available soil water (ASW) using a curvilinear relationship of the form $AETCF = \alpha - e^{(-\beta.ASW)}$ and generalisation from known relationships between eucalypt canopy conductance and soil water potential (Don White, personal communication). In the absence of reliable information on soil depth, a constant

depth of one metre was used for every plot. As a result differences between sites in this database analysis are predominantly due to climate, slope, aspect and soil texture.

A range of site drought-stress measures were tested including annual minima and means of soil water potential, ratio of available soil water to the maximum soil water store, the ratio of actual to potential evapotranspiration and the total annual run-off. Several measures (the mean of the ratio of actual to potential evapotranspiration for each month, the mean and the annual minimum of the ratio of available soil water to the maximum soil water store for each month and the mean of the soil water potential for each month) explained an equivalent amount of deviance when tested in a logistic regression context. The mean of the estimated soil water potential for each month ($\overline{\Psi}_s$) was preferred in subsequent analyses because it most clearly separated mixed-species and pure stands of the species, was conceptually simple and was easily related to soil water measures made in subsequent physiological experiments. This variable was used in a logistic regression model, assuming a binomial probability distribution, and using the logit link function. Up to the fourth power of this variable, and of annual minimum monthly temperature, were tested for significance using a forward stepwise selection procedure (PROC LOGISTIC, SAS Institute Inc. 1990d). The temperature term was used because of its significance in explaining the distribution of *E. obliqua*. Only variables significant at the $p<0.001$ level were included in predictive models.

5.2.2 Distribution of mature trees within mixed-species stands

A 750 m belt transect was randomly located in an area of even-aged *E. tenuiramis* and *E. obliqua* mixed-species forest close to the long-term growth plots described below. (Details of the transect are: 147° 49′ 10″ E, 42° 31′ 30″ S, altitude = 350-250 m, mean annual rainfall = 790 mm, mean annual evaporation = 960 mm.) *Eucalyptus pulchella, E. viminalis* and *E. globulus* occurred infrequently. The transect ran from the top of a hill and down the hill fall-line to a small drainage course below. The aspect was uniformly east and the slope varied between 0 and 10°. In subsequent analysis the effect of topography was assumed negligible and not considered further. The transect was divided into 75 contiguous 100 m^2 quadrats, and counts of the number of individuals of *E. tenuiramis* and *E. obliqua* made. In each plot three measurements of soil depth were made by driving a steel rod through the soil to the underlying bedrock or 100 cm depth, whichever came first. Soil pits indicated that this technique was impeded only by rocks greater than 30 cm in thickness, and that otherwise it accurately estimated the depth to bedrock.

The water-balance model described above was used to calculate $\overline{\Psi}_s$ for each plot, with all inputs except for soil depth held constant. This contrasts with the previous analysis where soil depth was held constant and climate, slope, aspect and soil texture varied. Consequently in this analysis soil depth, via its influence on available soil water, is the controlling factor of drought-stress. The soil type was a clay derived from Jurassic dolerite, and was classified as being in the 'clay' texture class for water-balance calculations. A Poisson regression with a log-link function (using

PROC GENMOD, SAS Institute Inc. 1993) was used to explain either the number of stems of *E. tenuiramis* or *E. obliqua* per plot. In each case variables related to $\overline{\Psi}_s$ (a polynomial up to order 3), the distance along the transect (again up to order 3) and the counts of the other species were tested for significance. The distance measures were included to test for patchiness unrelated to variation in soil depth.

5.2.3 Stand development

Five long-term growth plots, 400 m^2 in area, within a mixed-species forest were examined for changes in dominance during the first 15 years of stand development, following logging and regeneration (see Lockett & Candy 1984; Elliott *et al.* 1991). These plots were randomly located in an area of mixed-species forest in south-eastern Tasmania approximately 80 km north-east of Hobart (147° 48′ 30″ E, 42° 30′ 30″ S, altitude = 380-400 m). At age 15 years, one of these plots (SW30-4) was dominated exclusively by *E. tenuiramis*, one dominated exclusively by *E. obliqua* (SW30-1a), two dominated by *E. obliqua* but with small areas of *E. tenuiramis* dominance (SW30-1b & SW30-5), and one plot in which dominance was shared (SW30-3). In addition to *E. tenuiramis,* the ecologically analogous species *E. amygdalina* Labill. and *E. pulchella* Desf. occurred on the study plots. For this analysis these species have been pooled together into the series group *Piperitae* (Pryor & Johnson 1971) to increase sample numbers. For consistency, *E. obliqua* is referred to in this section of the analysis by its series title *Obliquae* (Pryor & Johnson 1971).

At the time of regeneration in 1979, the forest area was cleared-felled, burnt and resown with a seed mix representative of the prelogging species balance (77% *Obliquae,* 8% *Piperitae,* 15% *Symphyomyrtus* [being 11% *E. globulus* and 4% *E. viminalis*]). Each patch of soil, therefore, had an equal chance of receiving seed of a particular species, although any one patch was more likely to receive a seed of *E. obliqua* than any other species. The location, height, diameter and species of each eucalypt plant found on the plots were recorded intermittently (at year 1, 2, 3, 4, 6, 9, 12 & 15) until 1994. For each plot, soil depth above bedrock was measured on a 2×2 metre grid (i.e., 100 locations per plot) using the technique outlined above. At each location, up to three depth readings were taken. Each tree on each plot was assigned a soil depth by interpolating from a smoothed surface generated from the regular grid of points using a bivariate spline function (with PROC G3GRID, SAS Institute Inc. 1988). After checking the normality of data, analysis of variance was used to test for differences between the series *Piperitae* and *Obliquae* in the soil depth associated with the tallest 25% of the trees on each plot (using PROC GLM, SAS Institute Inc. 1990d). Where all plots were considered together, plots were treated as blocks in the analysis. The soil depths were used in estimating $\overline{\Psi}_s$, following the manner described above, for comparison with other data.

5.2.4 Species' performance on an artificial soil moisture gradient

Gradients in soil depth were created in six boxes of 3 m length by 0.3 m width which were covered with a transparent roof that protruded well beyond the edges of each box. Air flowed freely below this roof and conditions in the box were the same as ambient summer conditions in Hobart where humidity typically varies between 30 and 70% and mean daily maximum temperature is close to 20°C. The bottom of each box sloped from 15 cm deep at the shallow end to 50 cm at the deep end with the bottom of each box made of two layers of woven nylon mesh over gravel that allowed water to drain but constrained plant roots. Each box was filled with an organic mulch. The moisture release properties of this soil were determined using the filter paper method (Hamblin 1981).

In each box 19 rows were planted at 15 cm intervals. Each row contained two glasshouse-raised seedlings (≈ 20 cm in height), originating from seed collected at the transect site. Within rows, plants were separated from each other and the edge of the box by 10 cm. Two boxes were planted exclusively with seedlings of *E. tenuiramis*, two exclusively with seedlings of *E. obliqua* and two with one seedling of *E. tenuiramis* and one seedling of *E. obliqua* per row. The height of each seedling was measured immediately after planting. Plants were grown for four months under a regime of twice daily watering and twice weekly addition of complete nutrient solution. At this time plant roots occupied the total soil volume and there was no significant difference in the height of plants from differing positions on the soil depth gradient. Watering was then withheld for one month, the soil was raised to field capacity, and watering was again withheld for a further month.

At the end of this second drought-stress cycle, an open gas exchange system incorporating an Analytic Development Company Mk 2 infra-red gas analyser was used to measure net CO_2 and water vapour exchange by foliage. A lamp containing four 150 W Wotan xenon quartz globes was used to maintain an incident photon flux density at the leaf surface of 800 μmol m^{-2} s^{-1}. Air at ambient CO_2 concentration was pumped through the leaf chamber (cross-sectional area 6.25 cm^2) and the reference line at a rate of 400 ml min^{-1}. Measurements were taken between 10:00 and 14:00 hours, with replicates of each treatment selected randomly. The photosynthetic rate of three leaves on each plant was measured. Leaves were selected from nodes 4, 5 and 6 from the plant apex. The average of these three measures was used in subsequent analysis.

The following day, predawn water potentials of lateral shoots of all plants in rows 2, 10 and 18 were measured using a PMS pressure bomb. At the same time, 10 cm diameter soil cores to the full soil depth were taken between each plant pair, weighed, oven-dried and reweighed to calculate the relative water content at different positions along the soil depth gradient. This was transformed to soil water potential using the moisture release curve for the organic mulch soil to enable comparison with $\overline{\Psi}_s$ used in the geographic database study. At this stage, plants were rewatered and kept wet at field capacity for three days. The photosynthetic rate of all plants was

then remeasured. The heights of all plants were then remeasured and the above-ground biomass was harvested. Each harvested plant was divided into leaf and stem material, and dried at 40°C in a drying room. Sub-samples were oven-dried at 80°C and a correction factor applied to total samples. Sub-samples from each tree in rows 2, 6, 10, 14 and 18 were taken and a leaf area-to-weight ratio relationship was developed. A number of plants were excavated and it was observed that the plant roots were concentrated between one row above and one row below their planted position. However, intermingling of roots between neighbouring plants and the in-growth of roots into the organic mulch prevented root biomass assessment.

The canopies of plants in the gradient experiment were too small for the development of different leaf types (i.e. sun and shade leaves). All leaves were of approximately similar age and there was only limited self-shading. In scaling up from the leaf processes to an estimate of canopy productivity, the simple approach was adopted of assigning the average rate of photosynthesis derived from the measurement of three leaves in each plant to the total leaf area. This is clearly an over-estimation, but nevertheless may still provide a useful relative measure of potential canopy carbon uptake.

Differences along the artificial gradient in photosynthetic performance and potential canopy carbon uptake to soil relative water content were assessed by fitting a Gompertz curve to the data. This function, of the form $y = \alpha.\exp[-\exp(\beta-\delta.X)]$, fits a sigmoidal curve which can be asymmetric around the point of inflection (Ratkowsky 1990). The function was fitted to the data and confidence limits for parameters and the fitted curve were calculated using the SAS non-linear regression procedure NLIN set to the MARQUARDT iterative method (SAS Institute Inc. 1989).

5.3 Results

5.3.1 Database analysis of co-occurrence

Both *E. tenuiramis* and *E. obliqua* commonly form mixed-species stands with other members of the subgenus *Monocalyptus* (Fig. 5.1). *Eucalyptus tenuiramis* and *E. obliqua* occur in pure stands (without other *Monocalyptus* species) in 23% and 56% of occurrences respectively. Of all occurrences of *E. tenuiramis,* 48% are co-occurrences with *E. obliqua.* Of these stands, 40% were dominated by *E. tenuiramis* and 60% by *E. obliqua* (or 46% and 54% respectively if stands containing other species of the subgenus are excluded).

Logistic regression gave a high degree of concordance with the data: a pair of input observations with different responses is said to be concordant if the larger response has a higher predicted event probability than the smaller response. In all models concordance was in excess of 80%. Discordance (the opposite of concordance) was between 15% and 20% for all models. The curves and equations, and the data grouped into categories and expressed as an observed probability of occurrence are given in Figure 5.2.

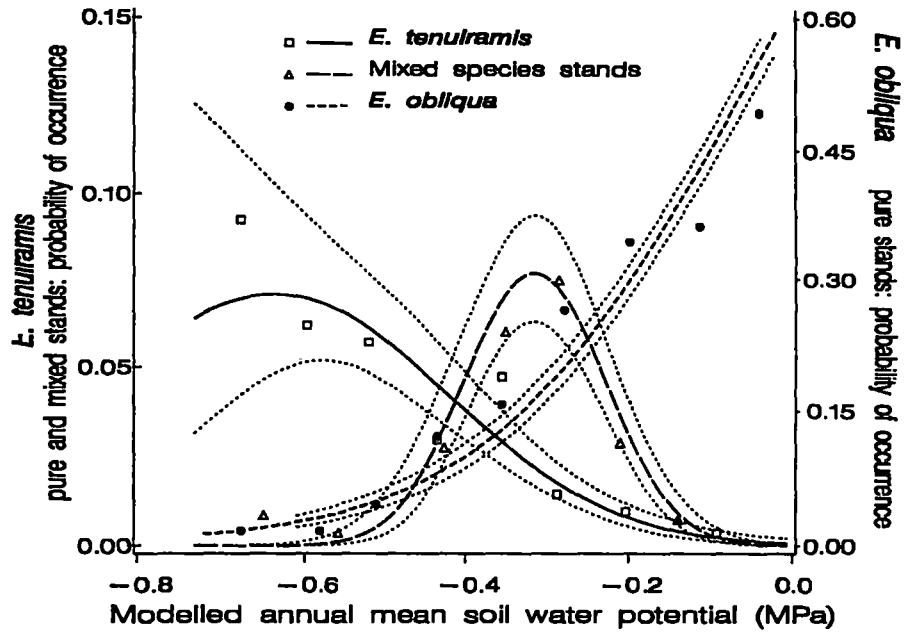


Figure 5.2 Actual and predicted probability of occurrence of pure and mixed-species stands of *E. tenuiramis* and *E. obliqua.* The *confidence intervals* are the 95% upper and lower limits of the mean predicted value. Note that *E. obliqua* is scaled on the *right-hand y-axis* and the *E. tenuiramis* and mixed-species stands on the *left-hand y-axis*. The linear predictors, η, for the logistic regression functions are: *E. obliqua* pure stands, $\eta = 4.7–6.7\,\overline{\Psi}_s–2.8\,MT+0.30\,MT^2$, where $\overline{\Psi}_s$ = estimated annual mean soil water potential (MPa), and MT = annual minimum monthly temperature (°C); mixed-species stands, $\eta = 9.2+45.2\,\overline{\Psi}_s+71.8\,\overline{\Psi}_s^2$; and *E. tenuiramis* pure stands, $\eta = 7.4+15.1\,\overline{\Psi}_s+11.9\,\overline{\Psi}_s^2$. In calculating the probability of *E.* obliqua, MT was set at 2.5°C. The actual data is grouped into ten 0.08 MPa classes and plotted at the class mean, but it should be noted that models were derived from continuous binomial data.

There was clear evidence that mixed-species stands occurred in environments intermediate to those environments occupied by pure stands of either *E. tenuiramis* or *E. obliqua*. On the wettest plots ($\overline{\Psi}_s > –0.1$ MPa), only *E. obliqua* pure stands were predicted to occur, and it was in these situations that the species had its highest probability of occurrence. *Eucalyptus obliqua* continued to have a higher probability of occurrence than mixed-species stands or pure stands of *E. tenuiramis* down to a $\overline{\Psi}_s$ of –0.5 MPa, albeit at a greatly reduced probability compared to the wettest plots in the dataset. By contrast the highest predicted probability of occurrence of *E. tenuiramis* was on the drier plots, although its probability of occurrence was predicted to decline when $\overline{\Psi}_s$ fell below –0.64 MPa. Despite its infrequent occurrence in the landscape compared to *E. obliqua*, *E. tenuiramis* was nevertheless more likely ($p < 0.05$) to occur on plots where the $\overline{\Psi}_s$ was below –0.5 MPa. Mixed-species stands of the two species were most likely to occur on plots with a $\overline{\Psi}_s$ of –0.3 MPa. Mixed-species stands of *E. obliqua* and *E. tenuiramis* were more likely to occur than pure stands of *E. tenuiramis* when $\overline{\Psi}_s$ was above –0.4 MPa ($p < 0.05$). Nevertheless mixed-species stands were less likely to occur than pure stands of *E. obliqua* at any level of $\overline{\Psi}_s$.

5.3.2 Distribution of mature trees within mixed-species stands

The mixed-species stand transect (Fig. 5.3) indicated a significant relationship ($p < 0.001$) between the number of individuals of either species in a 100 m^2 plot and the variation $\overline{\Psi}_s$ as influenced by the depth of soil above bedrock (Fig. 5.4). Neither the inclusion of spatial co-ordinates (y, y^2 and y^3, where y is the distance along the transect), nor the abundance of the other species (*E. obliqua* in the case of the predictive model of *E. tenuiramis*, or *E. tenuiramis* in the case of the predictive model for *E. obliqua*) significantly improved the explanatory power of the model ($p > 0.05$ in all cases). The presence of *E. tenuiramis* and *E. obliqua* (i.e. the predicted number of individuals per plot was greater than 0.5) were accurately predicted 87% and 95% of the time respectively. They were predicted to occur when they in fact were absent 15% and 13% of the time respectively. *Eucalyptus tenuiramis* and *E. obliqua* co-occurred on 9 plots, too few to model. The observed frequencies of these mixed-species plots, and data for plots with only one of the species, are given in Figure 5.4.

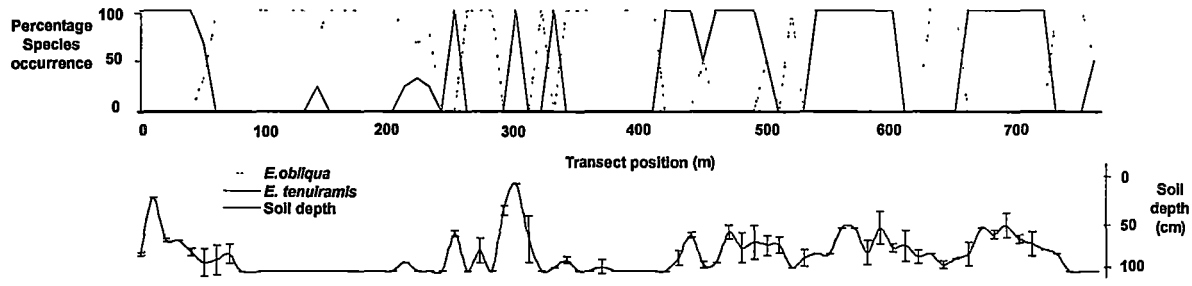


Figure 5.3. Diagrammatic representation of the transect through a mature mixed-species forest showing the relationship between soil depth and the relative abundance of *E. obliqua* and *E. tenuiramis* on contiguous 100 m^2 quadrats. *Error bars* are the 95% confidence interval of the mean soil depth of each plot.

The abundance of *E. tenuiramis* on the logarithmic scale was correlated to a quadratic function of $\overline{\Psi}_s$, with peak abundance predicted on plots in which the $\overline{\Psi}_s$ was –0.6 MPa or a soil depth of 50 cm (Fig. 5.4). The model was not significantly improved by the addition of higher order powers of $\overline{\Psi}_s$. It was predicted that individuals of *E. tenuiramis* were unlikely (i.e., < 0.5 predicted individuals per plot) to be found on plots when the $\overline{\Psi}_s$ was less than –0.75 MPa or above –0.45 MPa (i.e. soils shallower than 10 cm or deeper than 85 cm). The predicted abundance of *E. obliqua* increased linearly on the logarithmic scale with $\overline{\Psi}_s$ up to –0.4 MPa or a soil depth of 100 cm. Plots with a predicted $\overline{\Psi}_s$ of below –0.45 MPa were predicted to be unlikely to sustain mature individuals of *E. obliqua*.

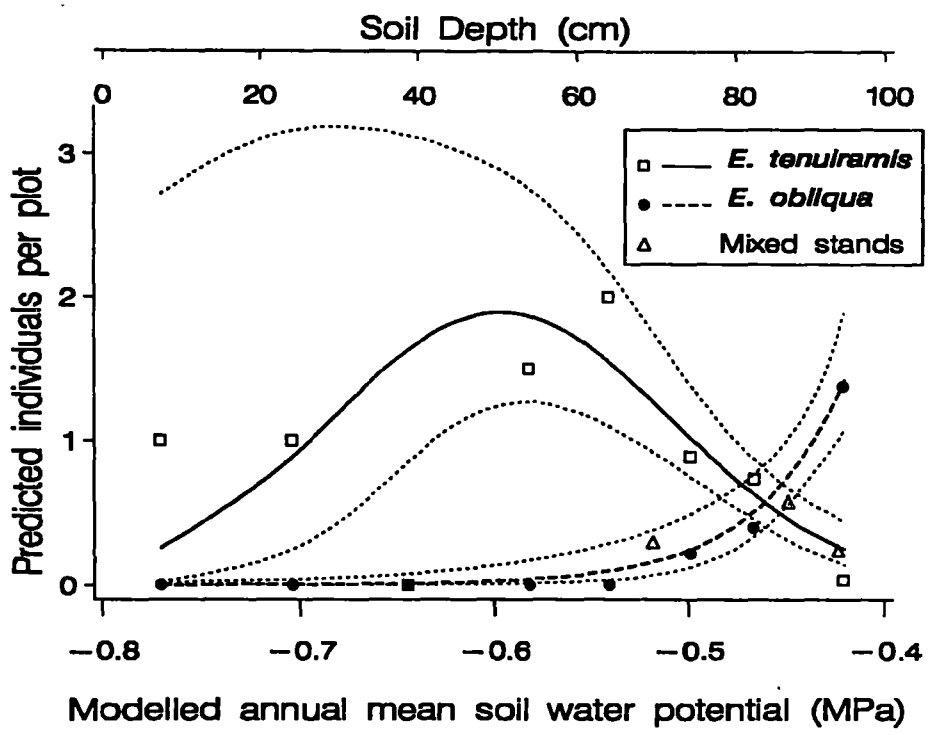


Figure 5.4 Predicted abundance of *E. obliqua* and *E. tenuiramis* and the estimated annual mean soil water potential ($\overline{\Psi}_s$) for soils of differing depth from 78 plots located in a mature mixed-species forest. The predicted number of individuals of each of *E. tenuiramis* and *E. obliqua* is given by TEN = exp(–22.54–65.32 $\overline{\Psi}_s^2$+77.84 $\overline{\Psi}_s$ and OBL = exp(9.65+22.17 $\overline{\Psi}_s$), respectively. The *confidence intervals* are the 95% upper and lower limits of each predicted value. The observed mean number of individuals for eight, 0.04 MPa classes are plotted at class means, but it should be noted that models were derived from continuous data.

5.3.3 Stand development

The initial distribution of seedlings from the series *Obliquae* and *Piperitae* was independent of soil depth and the plot $\overline{\Psi}_s$ ($p > 0.1$). The distribution of all surviving stems after 15 years was similarly independent of soil depth ($p > 0.1$), reflecting the low overall rates of mortality. However, if the tallest 25% of trees on each plot (referred to as the dominant trees), or any percentile above this, was considered alone, significant differences ($p < 0.05$) between the series were evident during stand development (Table 5.1). In the first year after seedling regeneration, there was no difference in the average soil depth associated with individuals from either series (*Obliquae* soil depth = 85 cm; *Piperitae* soil depth = 81 cm; $p > 0.05$). By 15 years of age, dominance by *Obliquae* was associated with deeper soil than was dominance by species in the series *Piperitae* (mean for all plots of 89.5 cm vs. 76.3 cm; $p < 0.05$). At this age there were significant differences in the soil depth associated with dominance by species in each series in three of the plots (SW30-1b, SW30-3 and SW30-5). On these plots, the differences emerged early in stand development some time in the first three years when trees were between 50 and 100 cm in height. On the remaining two plots, SW30-1a and SW30-4, all dominants at age 15 years were from the series *Obliquae* and *Piperitae* respectively.

Table 5.1 Tests for significant differences in the soil depth and the annual mean soil water potential associated with dominant individuals (the tallest 25%) of the *Eucalyptus* series *Obliquae* and *Piperitae* at various ages following clearfelling, burning and aerial sowing on 400 m^2 plots. Soil depths and the modelled annual mean soil water potentials, $\overline{\Psi}_s$, are derived from a systematic grid of 100 points per plot.

Plot	SW30-1a	SW30-1b	SW30-3	SW30-4	SW30-5	All plots
	Average plot soil depth (m) with the standard error in parentheses					
	85.3 (7.6)	85.2 (22.4)	78.7 (15.1)	70.8 (18.8)	95.4 (10.8)	
			$\overline{\Psi}_s$ (MPa)			
Age (years)	–0.46	–0.46	–0.48	–0.51	–0.43	
1	ns	ns	ns	ns	ns	ns
2	ns	**	ns	ns	ns	**
3	ns	**	**	ns	**	**
4	ns	**	**	ns	**	**
6	ns	**	**	ns	**	**
9	ns	**	**	ns	**	**
12	**	**	**	ns	**	**
15	all *Obliquae*	**	**	all *Piperitae*	**	**

ns = ($p > 0.05$); ** = ($p < 0.05$);

The total dominance of SW30-4 by *Piperitae* was despite the overwhelming weighting of *Obliquae* in the initial sowing mix. SW30-4 was the plot with the shallowest soil and had a $\overline{\Psi}_s$ of –0.51 MPa. SW30-1a was an unusual plot in that the soil depth was consistent within the plot with no marked demarcation into shallow or deep patches. The $\overline{\Psi}_s$ and soil depth associated with dominance by the series *Obliquae* and *Piperitae* were generally consistent with the observation for mature stands from the transect study: there appeared to be a critical soil-depth threshold, under the regional climate of these study sites, in the range 80 to 90 cm.

The separation of plot SW30-1b into patches of dominance by age 15 years is given in Figure 5.5 as an example. It can be seen clearly in this figure that dominant individuals of *E. tenuiramis*, from the series *Piperitae,* are associated with the shallow soils on one side of the plot despite the fact that seedlings are well scattered throughout the plot. *Eucalyptus obliqua* seedlings grew poorly in this patch.

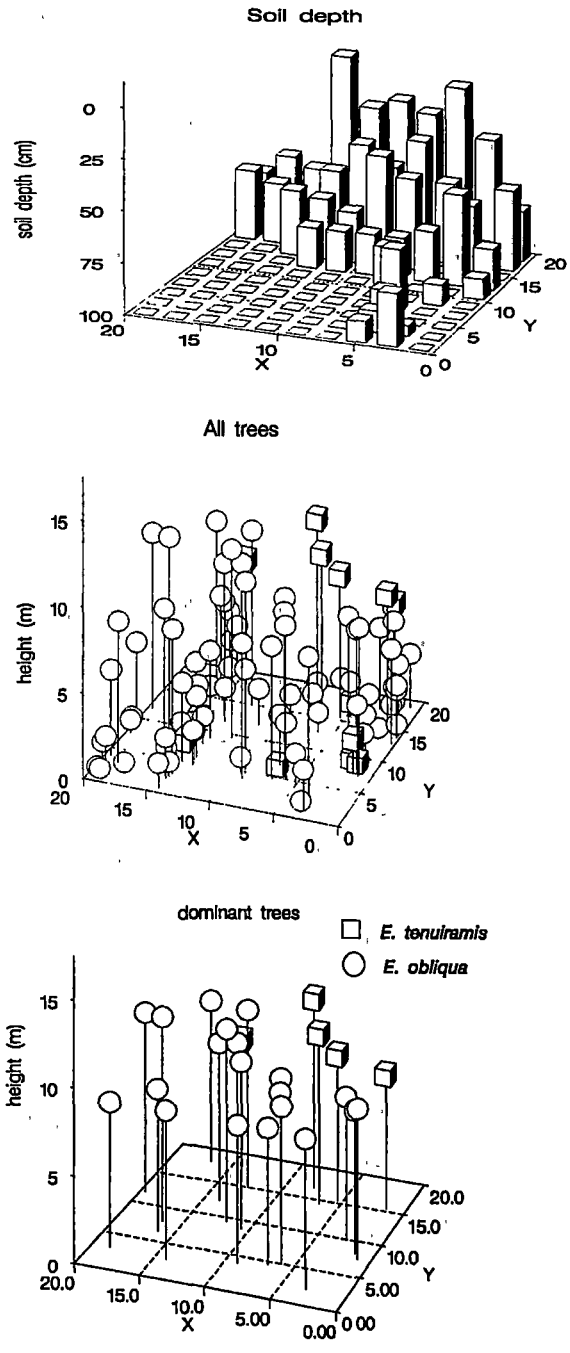


Figure 5.5 The soil depth (*tall bars* indicate shallow soil) and location and height of all trees, and the dominant trees only, on one of the 400 m^2 long-term growth plots, SW30-1b, 15 years after clearfelling and artificial sowing.

5.3.4 Physiological responses on an artificial water supply gradient

Following the completion of the second drought cycle, significant differences in soil water deficit were measured along the gradient. Soil at the shallow end had a relative water content of between 0.2 and 0.3 g g^{-1}, indicating a soil water potential of between –0.25 and –3.0 MPa. Soils at the deeper end of the gradient remained near to field capacity for the duration of the drying cycles (> –0.01 MPa).

These differences were reflected in the height increment of plants over the drought period, most particularly among *E. obliqua* plants (Fig. 5.6a). The height increment of *E. obliqua* increased linearly with soil relative water content along the soil depth gradient, even when this was

reflected in only slight changes in soil water potential. By contrast the response of *E. tenuiramis* exhibited a threshold response, with a critical value for soil relative water content of 0.4 g g^{-1} (equivalent to –0.05 MPa), below which the response was variable but generally sub-optimal, and above which it was variable but generally exceeded 25 cm. At the shallow-soil end of the gradient, the growth increment of both species was low and not significantly different. In the region of 0.30 to 0.45 g g^{-1} (–0.05 to –3.0 MPa) height increment of *E. tenuiramis* was significantly greater ($p < 0.05$). At the deep-soil end of the gradient the height increment of *E. obliqua* exceeded that of *E. tenuiramis*. Plants of both species performed similarly whether grown in monoculture or mixed culture. This was the case for all performance indicators tested, and subsequent results deal only with species comparisons.

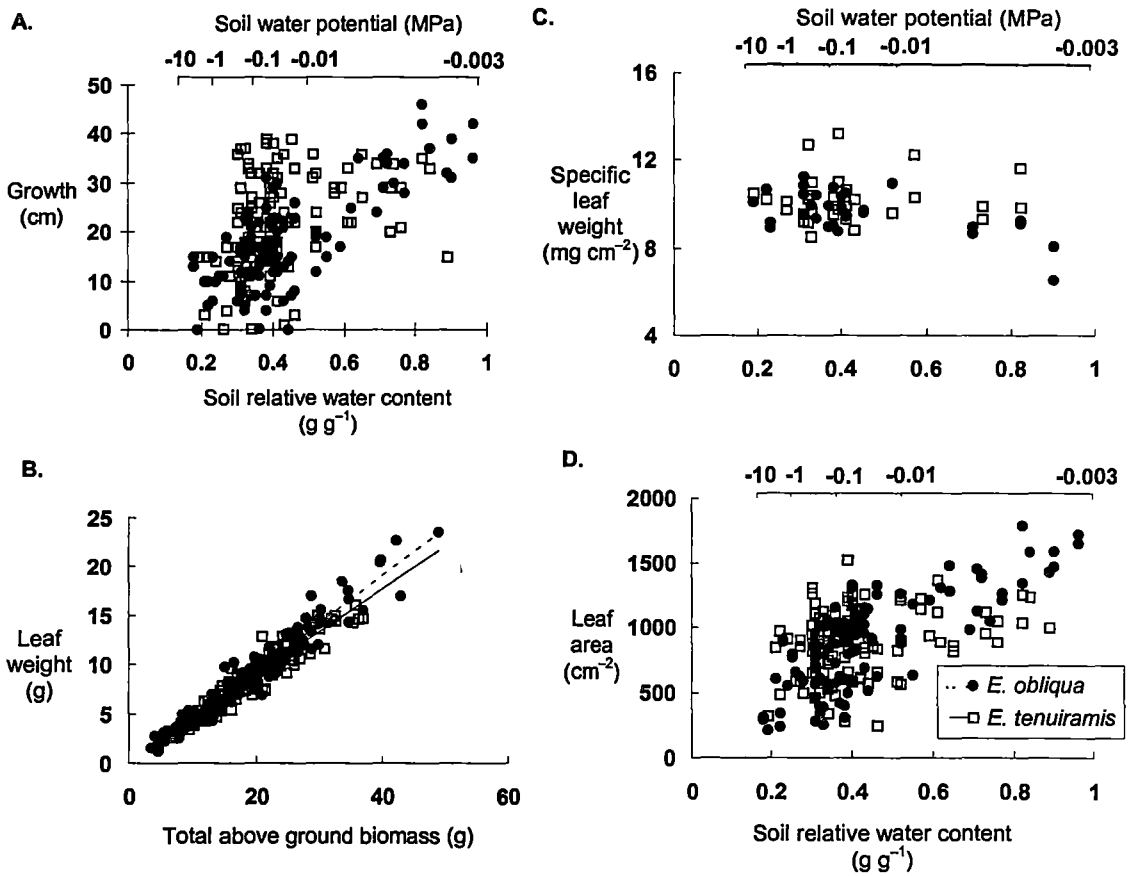


Figure 5.6 Growth and biomass relationships of seedlings from the artificial soil water supply gradient: A. the height increment over two droughting cycles of seedlings of *E. obliqua* and *E. tenuiramis* growing at different locations on an artificial soil water supply gradient; B. the proportion of above-ground biomass allocated to leaves by *E. obliqua* and *E. tenuiramis*; C. changes in specific leaf weight by the two species when grown under differing levels of soil moisture stress; D. the total leaf area per plant when grown under differing levels of drought-stress. The soil relative water content and the soil water potentials represent the maximum drought-stress induced.

In both species leaf weight was linearly related to total above-ground biomass (Fig. 5.6b). The proportion of biomass in leaves of *E. obliqua* (i.e. the slope of the line relating leaf weight to total above ground biomass) was significantly greater than for *E. tenuiramis* ($p > 0.05$), although the difference was slight (48% as opposed to 45%). Specific leaf weight was unrelated to the position on our imposed water supply gradient for the *E. tenuiramis* plants, but there was a trend

for leaves from *E. obliqua* plants growing on the deep-soil end of the gradient to be thinner (Fig. 5.6c). The additive effect of these two relationships, and the linear relationship between growth and drought-stress indicated above for *E. obliqua*, meant that at the deep-soil end of the gradient *E. obliqua* plants possessed a greater leaf area than did *E. tenuiramis* plants (Fig. 5.6d).

The effects of soil water content were reflected in pre-dawn leaf water potential of plants (Fig. 5.7). On the day photosynthetic measurements were made, the plants at the shallow-soil end of the soil depth gradient had pre-dawn leaf water potentials of below –1.5 MPa, while plants at the deep-soil end of the gradient appeared unstressed. There was no significant difference between the species in pre-dawn leaf water potential.

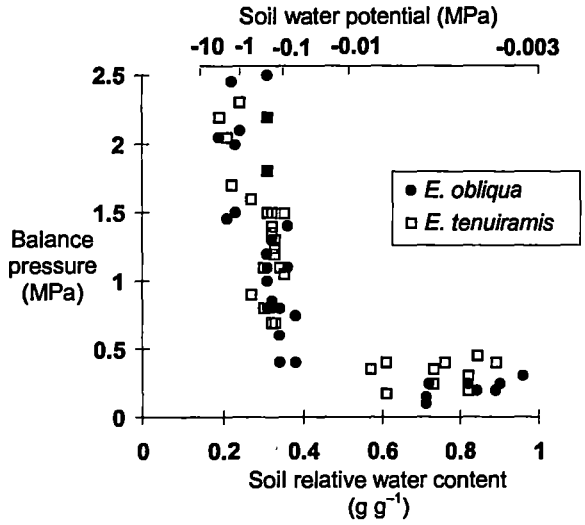


Figure 5.7 Predawn leaf water potentials of seedlings of both species growing at different locations on an artificial soil water supply gradient on the day CO_2 and H_2O flux rates were measured. The predawn leaf water potential is estimated as the negative of the balance pressure.

The rate of light-saturated, net photosynthesis at ambient levels of CO_2 (A_{max}) in both species declined with decreasing soil relative water content (Fig. 5.8). At both low and high levels of drought-stress there was no significant difference between species ($p > 0.05$), however at soil relative water contents of between 0.3 and 0.5 g g^{-1} (–0.25 and –0.02 MPa) the A_{max} of *E. tenuiramis* was significantly higher ($p < 0.05$). When drought-stress was removed by watering for three consecutive days, no significant differences were found in the A_{max} of seedlings irrespective of species or position on the soil depth gradients ($p > 0.05$), indicating that other factors associated with soil depth or plant position were not significant in this outcome.

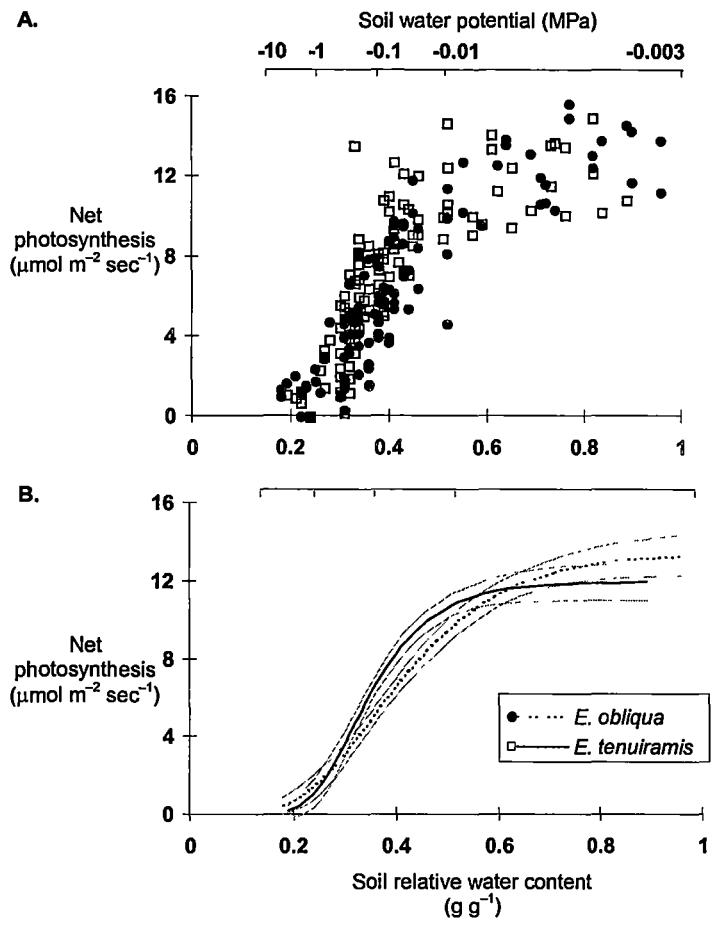


Figure 5.8 Net photosynthesis of *E. obliqua* and *E. tenuiramis* growing at different locations on an artificial soil water supply gradient: A. the measurements; B. fitted Gompertz curves and the 95% confidence interval of the relationships.

There were significant differences between the species in the response of stomatal conductance (*g*) to soil water content (Fig. 5.9A). These differences were consistent with the observed differences in assimilation rate; as with A_{max}, the stomatal conductance of *E. tenuiramis* plants at soil relative water contents of between 0.3 to 0.5 g g^{-1} was significantly higher ($p < 0.05$). There was, however, no difference between the species in the instantaneous water use efficiency (Fig. 5.9B). At soil relative water contents above 0.40 g g^{-1} the instantaneous water-use efficiency (calculated as the ratio of *A/g* after Osmond *et al.* 1980) for both species was approximately 1.75 to 2.0 μmol CO_2 fixed for every mole of H_2O transpired. At greater levels of drought-stress, water use efficiency declined linearly for both species at a similar rate.

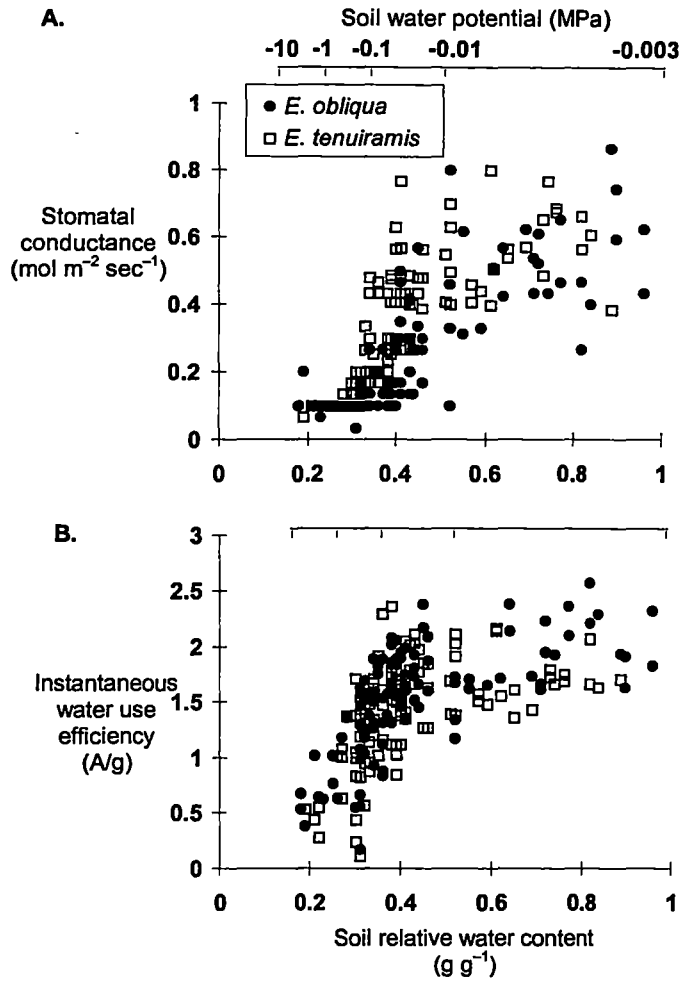


Figure 5.9 Stomatal conductance (A) and instantaneous water use efficiency (B) of *E. obliqua* and *E. tenuiramis* growing at different locations on an artificial soil water supply gradient.

The combined effects of drought-stress on A_{max} and the leaf area production of plants suggests that at low levels of drought-stress *E. obliqua* has a greater ability to fix carbon than does *E. tenuiramis* (Fig. 5.10). At intermediate levels of drought-stress the higher A_{max} of *E. tenuiramis* and the equivalent leaf area of the species indicates that *E. tenuiramis* may have a slight carbon uptake advantage. At high levels of drought-stress the low A_{max} of both species results in similar potential rates of carbon fixation per plant and generally poor growth.

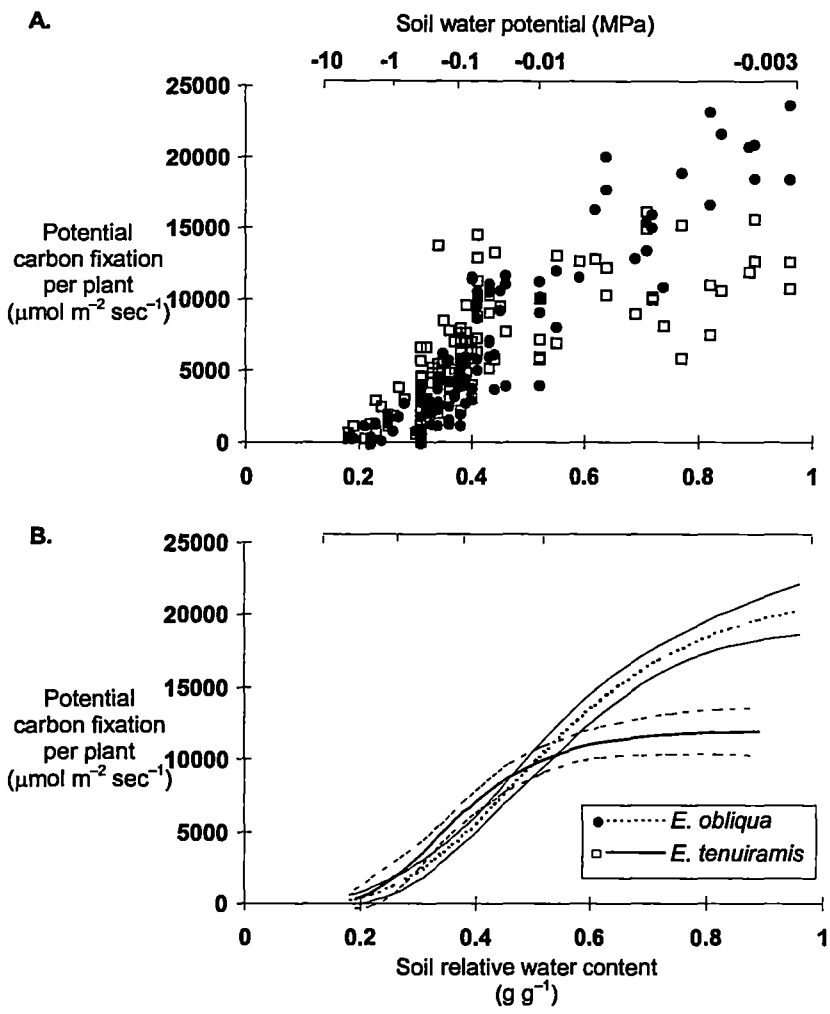


Figure 5.10 Estimated total canopy carbon fixation per plant of *E. obliqua* and *E. tenuiramis* growing at different locations on an artificial soil water supply gradient: A. the measurements; B. fitted Gompertz curves and the 95% confidence interval of the relationships.

5.4 Discussion

The occurrence of *E. tenuiramis* and *E. obliqua* in mixed-species or pure stands was correlated with site water supply at the broad geographical scale, largely through the influence of regional climate. Observed at a finer scale, mixed-species stands were composed of micro-stands. The distribution of these micro-stands corresponded to local variation in soil depth that influenced the water supply of patches. At both the broad geographical scale and the local scale of study, drought stress was the determinant of species' occurrence and abundance, although climate was the predominant discriminating factor in the former and soil depth in the latter. At both scales of study similar values of drought-stress ($\overline{\Psi}_s$) were associated with the peak probability of occurrence of the two species. A similar value of drought stress was also associated with the changeover of species at the local scale and the peak probability of mixed-species stands at the broad geographical scale.

The regional database study predicts that mixed-species stands are likely to occur over a broader range of $\overline{\Psi}_s$ than is observed when local variation in soil depth alone is the cause of variation in annual mean soil water potential. This difference is probably a result of between-plot variation in soil depth that was not considered in the database study in which a constant depth of one metre

was used. This will have resulted in an unknown error in the estimation of annual mean soil water potential. Additionally, the larger plot-size in the database examination (0.1 to 0.3 ha vs. 0.01 ha) probably resulted in more within-plot variation in soil depth. Consequently micro-stands of each of the two species (which the transect indicates can be less than 30 m in diameter) may have been classified as mixed-species stands.

This study indicates that the spatial shift from stands dominated by mature individuals of *E. tenuiramis* to stands dominated by *E. obliqua* occurs at an annual mean soil water potential of around –0.45 to –0.5 MPa. At the transect site and the nearby long-term growth plots this value corresponded to a soil depth of approximately 85 cm. Variation in soil depth above and below this range influenced the abundance of *E. obliqua* and *E. tenuiramis* respectively, rather than species' presence or absence. This soil depth threshold responsible for species' changeover is obviously site-specific. It will be influenced by climate and by site factors that affect water inputs (such as topography and slope), factors that affect water losses (such as aspect and vegetation cover), and factors other than soil depth that affect soil water storage (such as rockiness and soil texture and structure). Consequently, at any one location it is the combination of the mean site climatic regime and the amplitude of variation in conditions at the local scale that determines vegetation composition.

The onset of patch differentiation and the development of micro-stands coincided in the field with the age at which plants were large enough to explore the full soil volume, and hence at a stage when the variation in soil depth could result in between-patch differences in soil water supply to plants. The root biomass of *Eucalyptus* typically equals, or exceeds, the above ground biomass (Mulligan & Sands 1988; Cromer & Jarvis 1990) and eucalypt tap roots can explore the soil to great depths (Carbon *et al.* 1980). In the first year of growth, seedlings in this study were unable to fully exploit the total soil volume, and all seedlings experienced a similar soil environment irrespective of the actual soil depth in their immediate vicinity. By age 2 or 3 years, when seedlings were between 50 and 100 cm tall, seedling roots were probably capable of exploring the soil to a depth of at least 100 cm. The environment that had been homogeneous at the seedling stage had potentially become heterogeneous at the sapling stage, and this had begun to be reflected in plant performance.

The artificial gradient experiment indicated that both species grew best under conditions of high water supply, and that under these conditions carbon uptake rates were similar. *Eucalyptus tenuiramis,* however, maintained higher rates of CO_2 fixation under greater levels of drought-stress than did *E. obliqua*. The ability of *E. tenuiramis* to maintain a higher rate of stomatal conductance at lower levels of pre-dawn leaf water potential than *E. obliqua*, and the absence of differences in instantaneous water use efficiency between the two species, suggests that the occurrence of *E. tenuiramis* on drier sites is related, at least in part, to its ability to tolerate cytoplasmic desiccation rather than an ability to restrict water loss. This appears to be a common

method of drought tolerance among eucalypts, both in the differentiation of species (Withers 1978; Sinclair 1980) and between ecotypes of the one species (Ladiges 1974, 1976).

Eucalyptus obliqua continued to increase its leaf area with increasing water supply after the response in *E. tenuiramis* reached a maximum level. This was the additive result of the differences between the species in above-ground carbon partitioning and the lower specific leaf area at high levels of water supply among *E. obliqua* plants. Consequently, the potential carbon uptake per *E. obliqua* plant, in the absence of other resource limitations, was significantly greater at high levels of water supply. As a result, the growth rate of *E. obliqua* continued to increase at resource levels for which *E. tenuiramis* growth had reached a maximal rate.

It is clear from a number of studies that the ecological success of a species is only partially related to carbon uptake at the leaf (Givinish 1986; Field 1988), and that the pattern of carbon allocation (among other things) can be a major determinant of plant performance at different levels of resource supply (Schulze 1986). Sinclair (1980) concluded that the greater sensitivity of *E. obliqua* to drought stress than co-occurring species, *E. fasciculosa* F.Muell. and *E. leucoxylon* F.Muell., was probably due to the depth or extent of the root systems. Similarly, Mooney *et al.* (1978) noted that while no trend in assimilation rates and the aridity of eucalypt habitat could be found, a consistent trend between the aridity of habitat and the allocation pattern to leaves in *Eucalyptus* existed. In addition, the study indicated that the expression of this difference in biomass partitioning can itself be dependent on the test conditions. In our study it was only at moderate to high levels of soil water supply that the trends observed by Mooney *et al.* (1978) were apparent.

The fundamental trade-off between rapid growth and tolerance of low soil water, implicit in the conflict between CO_2 uptake and water loss (Smith & Huston 1989), is highlighted by the performance of *E. obliqua* and *E. tenuiramis* over different portions of our artificial water supply gradient. The response of *E. obliqua* is typical of mesophytic tree species that conserve moisture during times of high drought-stress by partial stomatal closure, foregoing photosynthesis and growth as a means of tolerating moisture limitations (Hinckley *et al.* 1978; Burton & Bazzaz 1995). The plasticity of biomass allocation in response to the level of resources observed between the *E. obliqua* plants in this study is another characteristic of species from resource-rich (in this case at least relatively so) environments (Bloom *et al.* 1985; Smith & Huston 1989). This trait allows species like *E. obliqua* to dominate what becomes the limiting resource of light when released from other resource constraints. By comparison, *E. tenuiramis* possesses a low maximum rate of growth in the absence of resource limitations, a common trait among plants adapted to sites characterised by predictably low levels of resource supply (Chapin 1980; Bloom *et al.* 1985). Many plants adapted to low resource conditions are unable to continue to increase their growth as conditions become more favourable, perhaps because of genetically fixed traits such as high root:shoot ratios (Bloom *et al.* 1985).

The carbon acquisition and growth strategies of the two species of *Eucalyptus* studied give each a growth advantage on a portion of the water supply gradient. Consequently, a mosaic of species' dominance will occur in spatially heterogeneous landscapes where the environmental variability shifts conditions to favour one or other of the species. This situation is most likely to occur where environmental gradients are shallow and conditions are marginal between those which favour the development of pure stands of *E. tenuiramis* or *E. obliqua*. It is, therefore, reasonable to consider these mixed-species stands of *E. tenuiramis* and *E. obliqua* as ecotones on a water supply gradient, occupying an environmental space intermediate to that of pure stands of either species. Spatial heterogeneity in conditions that affect soil water supply acts within this ecotonal environment to provide the means for species' co-occurrence.

The responses of *E. obliqua* and *E. tenuiramis* observed in this study support a number of the propositions of the continuum concept as proposed by Austin & Smith (1989). The proposition that species' fundamental response curves will be such that in a particular portion of the resource space a species will have a relative performance advantage over other species was supported. Both species showed similar monotonic responses to soil water supply in leaf-area, net photosynthesis, growth and modelled canopy CO_2 uptake. However, as discussed above, slight differences in the form of the response lead to *E. obliqua* having a growth advantage at high levels of water supply and *E. tenuiramis* a growth advantage at intermediate levels of water supply that nevertheless would be severe in the Tasmanian landscape. At very low levels of water supply neither species appeared to have an advantage on the basis of the traits examined. However, differentiation might be apparent on other niche dimensions, such as the ability to survive very severe drought, or total carbon uptake may not compensate for maintenance respiration costs and in the long-term neither species would persist. Nevertheless, as suggested by Austin & Smith (1989), the physiological niche of each species explains their observed ecological niches in the field.

Finally, there is clear support for the hypothesis that mixed-species stands are an artifact of landscape heterogeneity. The perception of mixed-species stands of these two species appears to be a matter of measurement scale. This study suggests that the occurrence of mixed-species stands is a result of local-scale variation within an environment intermediate to those favoured by pure stands of either species. Examination of what appeared to be a mixed-species stand when considered at the broad geographic scale was in fact composed of micro-stands when examined at a finer scale.

5.5 Conclusions

These results demonstrate that regional databases of floristic inventories can be robust tools for exploratory analyses of plant distributions. A direct environmental gradient in water supply, even though derived from limited site information, provided a mechanism for comparison between different scales of observation and between experimental and database studies (e.g. Williams &

Eamus 1997). The complementary interpretation of ecological processes from patterns of plant species' distributions and physiological responses supports the continuum concept and confirms the importance of defining gradients in terms of environmental factors which directly influence plant physiological responses.

6. Productivity gradients: can physiological processes be used as a bioassay of plant responses for predicting species' distributions?

6.1 Introduction

This chapter continues the theme of exploring the potential for predicting *Eucalyptus* species' distributions with derived indices that are more proximal to plant physiological processes than their original estimates of climate or parent rock type. Unlike annual rainfall and evaporation (illustrated in Chapter 4), temperature and solar radiation are two environmental gradients that have direct physiological influences upon plant growth and reproduction (Austin & Smith 1989). All other factors being plentiful, the light quality, duration of sunlight and the day- and night-time temperature regimes predominantly influence net photosynthesis and respiration. Carbon gain (difference between photosynthesis and respiration plus tissue loss) enables the plant to direct energy resources toward growth, reproduction or other physiological processes that ensure survival during periods of stress.

These primary plant physiological responses have lead to the development of productivity indices which can be used as surrogate measures for predicting plant distributions (e.g. Emmingham 1978; Specht 1981, 1983; Wright 1983; Wylie & Currie 1993). An index of potential productivity represents a bioassay of site conditions in terms of plant responses. A bioassay index is useful because different sets of environmental factors can result in similar productive potentials. For example, the height of mature native forest vegetation has frequently been used as a predictor of potential productivity for forestry planning, even though tall forests predominate in a range of environments (e.g. Lawrence 1978; Scott & Kohl 1994). However, tall forests can grow very slowly, and may essentially have a low productivity, so these structural properties are usually standardised by permanent plot studies, monitoring the increment in growth of different aged stands (e.g. West *et al.* 1984; West 1995; West & Osler 1995).Nevertheless, the measurement of bioassay factors, such as above ground biomass or productivity, provides a rapid means of establishing accumulated differences between sites resulting from the interaction between plant responses and environment. This approach is justified while the partitioning of carbon gain to above ground components, per unit time, is generally proportional to total productivity (e.g. Walters *et al.* 1993).

Plant ecologists have also used above ground biomass as a site index for simplifying the complex of interactions when comparing and interpreting plant species or community responses (e.g. Grime 1977; Tilman 1988; Keddy 1991; Bazzaz 1991). For example, a model of 'centrifugal organisation' of plant communities was developed by Keddy and co-workers to integrate pattern and process in ecological research (Moore *et al.* 1989; Keddy 1990; Keddy & McLellan 1990;

Wisheu & Keddy 1992). Their idea of plant community organisation was based on the observation that different combinations of environment can be biological equivalents, although this may not always be the situation. This approach has been criticised (Austin & Smith 1989; Austin & Gaywood 1994) on the basis that species' responses are individualistic. Even though the physiological processes leading to the differences in plant distribution responses are unclear (e.g. due to the confounding effects of competition, disturbance history and the seasonal or inter-annual extremes of climate), the standardisation of environmental factors by collective vegetation properties has application for the development of ecological theory.

A gradient that reflects the relative effects of stress and productivity is appealing as a surrogate for predicting plant distributions because it condenses the multivariate character of the environment into a single factor, effectively standardising the environment by a plant response (even though this may confound the effects for comparison between species). For example, temperature and light influence the growth of plants in a number of different ways. Apart from their direct effect on the rate of metabolism catalysis, cold temperatures in combination with high light intensity may induce photoinhibition, frost may disrupt cell membranes, and heat may denature enzymes (e.g. Berry & Bjorkman 1980; Kappen 1981; Ogren & Evans 1992; Eilers & Peeters 1993). As a result of adaptation to the different temperature regimes present in their environment, species may have a distinct environmental range, within which plant processes are optimised (e.g. Paton 1980; Criddle *et al.* 1996a, b). The accumulated response to such environments maybe skewed toward high temperatures and tailed toward low temperatures (e.g. Landsberg 1986; Austin 1992). Each physiological process (e.g. respiration, photosynthesis, cell elongation, flowering, seed maturation, germination) can therefore be characterised by an optimum response and limits for superoptimal and suboptimal conditions.

Temperature is also a cue for plant developmental processes. Cued responses are an adaptation to predictable, seasonal or inter-annual climatic regimes for minimising the damage to sensitive growing tissues and maximising potential growth. Since the seasonal regime of temperature and light are highly correlated, these two environmental factors interact with physiological processes. For example, the timing of flowering and the breaking of winter or summer dormancy patterns are probably determined by a combination of thermal and light regimes (e.g. Murray *et al.* 1994; Rundel *et al.* 1995). The whole plant response is therefore determined by the physiological summation of these processes allowing distribution limits to be directly correlated with the thermal tolerance of key physiological processes (e.g. Woodward & Williams 1987).

A large body of information about plant physiological processes exists (e.g. Slatyer & Ferrar 1978; Paton 1983, 1987; Kirshbaum & Farquar 1984; Kriedemann 1986; Leuning 1995). These studies have contributed to the information needed to simulate the physiological response of individual species (e.g. Whitehead *et al.* 1992; Aber & Federer 1992; Bonan & Sirois 1992; Bonan 1993; Vaisanen *et al.* 1994; Nikolov *et al.* 1995; Aber *et al.* 1995; Wang & Polglase 1995; Battaglia & Sands 1997). The accumulation of such information for a number of species

has enabled the development of theoretical models for physiological processes that are applicable to plant functional types (e.g. Sall & Pettersson 1994; Haxeltine & Prentice 1996b). These models of the plant fundamental niche could be used to redefine environmental factors in terms of physiological gradients. The development of a relationship between plant species' distributions and their environment, including information about their specific physiological responses, might be expected to clarify the relative importance of fundamental and realised niche processes. A more direct comparison of the inter-relatedness of response between species or other taxonomic and ecological groupings may also be possible (e.g. Emmingham 1978; Osmond *et al.* 1987; Neilson 1993; Cao *et al.* 1995; Brzeziecki *et al.* 1995; Mellilo *et al.* 1995). Potentially, an exploration of the general physiological and ecological processes of a species' distributions may clarify theories of community assembly that derive from a simple observation of above-ground vegetation biomass.

In this chapter the responses of four Tasmanian *Eucalyptus* species (*E. obliqua, E. regnans, E. amygdalina, E. globulus*) are examined to test whether prediction is improved by recombining the gradients for light and temperature into productivity indices. In doing so, this study also aims to test the assertion of Austin & Gaywood (1994; see also Austin & Smith 1989) that *a priori* definition of productivity gradients is inappropriate for ecological analyses of plant distribution data. A number of different levels of indices are tested, from simple recombination of temperature (as growing degree days and stress degree days with consideration of day-length), to more complex, and perhaps individualistic, combinations as carbon uptake indices.

A simple approach to modelling canopy photosynthesis and carbon gain was developed for the generic physiological response of *Eucalyptus* species in cool temperate habitats. A generic response facilitates comparison of the ecological performances between species along the same gradient in productivity potential.

The univariate ecological responses of individual species from logistic regression analysis were compared between these process-oriented gradients, thermal sum indices, and direct gradients in light and temperature. Since the significance of these factors may alter in the multivariate case, these responses to light and temperature were also considered in the context of their mean annual conditions and with indicative variables for water and nutrients. The potential of these gradients for simplifying ecological response models and for clarifying plant distributions with respect to environmental productivity and stress were assessed.

6.2 Method

6.2.1 Ecological dataset

The objective of this analysis was to estimate photosynthetic productivity for the 15 640 observations in the ecological dataset for the occurrence of Tasmanian *Eucalyptus* species (previously described in Chapter 2 and 3). This dataset contains information for both species

presence and absence. The representation of species and environments in these data, and the appropriate data subsets for modelling species' distributions were defined in Chapters 2 and 3. Site locations (latitude, longitude and altitude) enabled long-term monthly averages of climate to be estimated from the process model, ESOCLIM (see McMahon *et al.* 1996). The climate variables relevant to the development of a gradient in temperature and light-dependent photosynthesis comprised long-term monthly averages for minimum and maximum temperature, flat-surface and cloud-terrain adjusted solar radiation. In the absence of genetic parameters for species' responses (photosynthetic optima, ranges and acclimation potential), an average response for the genus *Eucalyptus* was derived (M. Battaglia personal communication, after Battaglia *et al.* 1996; Battaglia & Sands 1997).

6.2.2 Environmental gradients in temperature

The long-term monthly averages for mean daily minimum and maximum temperature were used to define several types of environmental gradient that may contribute to plant physiological responses. Mean minimum temperatures were taken as a general record of the night-time minimum temperature and the mean maximum temperature was taken as indicative of the day-time maximum temperature. The seasonal variation in temperature is also highly correlated with the light regime and follows the changes in day length, that can be simply estimated for a site from its latitude and orientation to the sun (e.g. Iqbal 1983). Quasi-daily values for maximum and minimum temperatures were interpolated from the monthly estimates by the fit of a high-order polynomial to the middle day of each month. Diurnal temperature variation (Fig. 6.1) was subsequently calculated as sine curves from the timing of sunrise and sunset (M. Battaglia personal communication); assuming that the coldest time of the day was one hour before sunrise, and the warmest time of day was one hour after midday (Box 6.1).

Box 6.1 Calculation of thermal gradients from the long-term estimates for monthly mean daily maximum and minimum temperatures.

Quasi-daily values (where i = 1,2,3...365 days) are interpolated by a polynomial spline between the middle of each month.

Mean daily temperature, $\overline{T_i}$ is calculated

$$\overline{T_i} = \frac{(T_{\min_i} + T_{\max_i})}{2}$$

as the average of long-term monthly mean daily maximum and minimum temperatures ($T_{\max_i}$ & $T_{\min_i}$, °C).

The timing of sunrise and sunset, and therefore the day-length were calculated for the middle day of each month.

The diurnal temperature variation ($T(t)_i$) is inferred from quasi-daily values of long-term monthly mean daily maximum and minimum temperatures, where the minimum temperature ($T_{\min_i}$) is defined to occur one hour before sunrise ($t_{\min}$) and the maximum temperature ($T_{\max_i}$) at one hour after midday ($t_{\max}$). Temperature varies sinusoidally with time between these two extremes, but the diurnal asymmetry requires three steps, defined around the 24 hour time period of the day (t):

For the first time period: $t_{\min} <= t <= t_{\max}$,

$$a_1 = \frac{180}{(t_{\max} - t_{\min})}$$

$$b_1 = 180 \times \frac{t_{\min}}{(t_{\max} - t_{\min}) + 90}$$

For the second time period: $t_{\max} < t <= 24$ then

$$a_2 = \frac{180}{(t_{\max} - t_{\min} - 24)}$$

$$b_2 = 180 \times \frac{t_{\max}}{(t_{\max} - 24 - t_{\min}) - 90}$$

For the third time period: $24 < t < t_{\min}$,

$$a_3 = \frac{180}{(t_{\max} - t_{\min} - 24)}$$

$$b_3 = 180 \times \frac{t_{\max}}{(t_{\max} - 24 - t_{\min}) - 90}$$

Over the three time periods:

$$T(t)_i = \left(\frac{(T_{\min_i} + T_{\max_i})}{2} \times \frac{(T_{\min_i} - T_{\max_i})}{2} \right) \times \sin\left(\frac{\pi}{180} \times (a_j \times t - b_j) \right)$$

where i = 1,2,3...365 days, j = 1,2,3 time periods and t = 1,2,3..24 hours

Mean daily day-time temperature $\overline{T_{\text{day}_i}}$ is calculated from the diurnal temperature variation by taking the average between sunrise and sunset.

A thermal gradient for the annual duration and magnitude of temperatures potentially suitable for growth and photosynthesis (GDD, day°C) is calculated from the accumulation of diurnal temperatures above 0°C between sunrise and sunset. This gradient assumes that growth is dominated by photosynthesis.

For the hours: *sunrise* < t < *sunset*:

$GDD = \Sigma_i \max(0, T(t)_i)$ where i = 1,2,3...365 days.

A cold stress gradient for the annual duration and magnitude of temperatures that may cause stress or injury (SDD, day°C) is calculated from the accumulation of diurnal temperatures below 10°C; disregarding time of day. This gradient assumes that cold temperatures may injur all types of metabolic processes.

$SDD = \Sigma_i \min(10, T(t)_i)$ where i = 1,2,3...365 days.

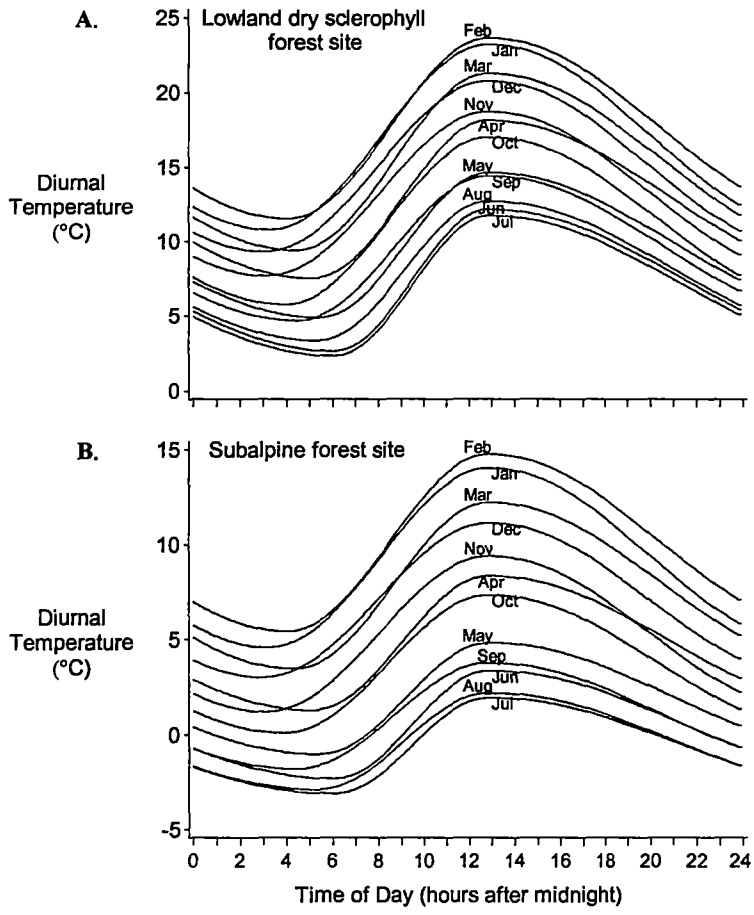


Figure 6.1 Examples of seasonal variation in diurnal temperature defined by an asymmetric sine interpolation of maximum and minimum monthly mean daily temperatures. Note the reversed positions of June and August between sites of contrasting thermal regime (different scales were used). Minimum temperatures were set for one hour before sunrise and maximum temperatures for one hour after noon.

Maximum and minimum estimates of average temperature may have differing significance for plant species' distributions. While the mean annual minimum or maximum temperatures may be important to the magnitude of a species' response, the absolute minimum and maximum temperatures reflect seasonal extremes, and are more likely to determine the limits to a species' distribution (e.g. Woodward & Williams 1987). Therefore, the three annual parameters (mean, minimum and maximum) for each monthly temperature variable (minimum and maximum) may need to be considered in ecological models of species' responses. The limiting physiological response for a species could then be inferred from the set of thermal variables selected in a correlative model of its ecological performance. The average daily, or day- and night-time temperatures can be estimated from the timing of sunrise and sunset and the model of diurnal temperature variation (Box 6.1).

Thermal sum variables can be estimated from the model of diurnal temperature variation as the average duration and intensity of conditions above or below a critical temperature. Two contrasting indices were defined: growing degree-days and stress degree-days (Box 6.1). The definition of critical temperature values required a consideration of the type of habitats represented by the study area. Eucalypt forest in Tasmania is evergreen and sclerophyllous vegetation. The warmer thermal regimes of lowland and coastal regions represent average winter minimum temperatures of about 5°C, in the transitional range between chilling sensitive to chilling resistant plants. In mountain and inland situations, the average winter minima

fall below freezing point where only species resistant to chilling may be expected to persist. Therefore, critical temperatures that were considered relevant to annual and seasonal estimates of the thermal sum indices were 10°C and 0°C. A critical temperature threshold of zero (freezing point) combined with day length was used to separate the growth response, and a critical temperature of 10°C (irrespective of the time of day) was used to separate the stress response. The overlap between these temperature thresholds reflects the potential for growth or stress in the 0-10°C range, which may differentiate species' distributions. For example, a cold tolerant species such as *E. delegatensis* may be able to acclimate for growth in temperatures between 0°C and 10°C. However, a lowland species such as *E. obliqua* may be prone to cold stress, or may not be able to compete with the faster growth rates of cold tolerant species where temperatures are sustained below 10°C.

Differences in diurnal temperature ranges (difference between maximum and minimum temperatures) between sites are frequently related to levels of continentality. Temperature ranges at coastal sites tend to be smaller due to the buffering effect of the ocean. Inland sites typically have wide temperature ranges, both seasonally and diurnally. These temperature effects represent an important component of the adaptive environment of a species. Apart from the direct effect of temperature variations on the acclimation and tolerance requirements of a plant, temperature influences evaporative demand through vapour pressure deficits. Vapour pressure deficits are an important measure of the drying power of air and reflect the difference between the saturation water vapour pressure and the partial pressure of water vapour in unsaturated air. The approximations to daily vapour pressure deficit from the long-term monthly averages for mean daily minimum and maximum temperatures, as suggested by Landsberg (1986), were therefore calculated (see Box 6.2). Estimates of vapour pressure deficit from diurnal temperature ranges therefore have a direct relationship with physiological processes and may be more significant in explaining species' distributions than simple calculation of the unweighted difference between daily minimum and maximum temperatures. More specific aspects of water balance can be determined by considering the role of rainfall and evaporation interacting with the vegetation and its substrate (see Chapter 4).

Box 6.2 Calculating mean daily maximum diurnal vapour pressure deficits from maximum and minimum daily temperature (adapted from Landsberg 1986).

Saturation water vapour pressure (kPa):

$$e_s(T) = 0.61078 \times \exp^{[17.269 \times T/(T+237.3)]}$$

is the partial pressure of water vapour in air at a particular temperature, T.

Partial pressure of water vapour in unsaturated air (kPa):

$$e_a = e_s(T_w) - \gamma(T - T_w)$$

where T_w is the wet bulb temperature, $e_s(T_w)$ is saturation vapour pressure at T_w, and γ is the psychometric constant (0.066 kPa°C^{-1}).

Vapour pressure deficit, $D = e_s(T) - e_a$ can subsequently be calculated.

To obtain estimates of vapour pressure deficit, e_a is assumed constant throughout the day, with the dew point temperature equal to T_{min} (pre-dawn), i.e. $e_a = e_s(T_{min})$. For an estimate of temperature at any time, vapour pressure deficit can be calculated.

Therefore, on a daily basis, the maximum and minimum estimates for temperature can be used to approximate maximum vapour pressure deficits, where $D = e_s(T_{max}) - e_s(T_{min})$. The diurnal temperature ranges may be an indirect approximation of vapour pressure deficits.

The seasonal acclimation environment in which a species is situated may also have important implications for the trade-off in physiological processes necessary to maintain optimum rates of photosynthesis and carbon gain in the presence of other species. The rate of change in temperatures between seasons is accelerated during autumn and spring periods, particularly in the southern, cool temperate environments of Tasmania. Changes in day length reach a maximum at the equinox (late April and late September), and influence climatic factors such as wind and diurnal temperature ranges. Since average day-time (maximum) and night-time (minimum) temperatures may have different magnitudes of influence on plant photosynthesis and respiration, separate indices were estimated for the seasonal rates of change in temperature. The rate of change in minimum or maximum temperatures were estimated from the difference between successive quasi-daily values and averaged per month.

For each set of monthly variables thus defined, one annual (mean of the 12 months) and two seasonal (maximum and minimum of the twelve months) were calculated for subsequent testing of their significance as factors correlated with *Eucalyptus* species' distributions.

6.2.3 Eucalypt canopy productivity

Three gradients in potential site productivity were defined from a model of eucalypt forest canopy photosynthesis and leaf respiration (Battaglia & Sands 1997). An estimate of the light-saturated photosynthetic rate describes a physiological response to the thermal environment, and canopy photosynthetic productivity integrates this temperature response with the light regime. Canopy carbon gain was subsequently estimated from the difference between canopy photosynthetic productivity and the losses due to leaf respiration. For the purpose of defining a gradient in relative site productivity, the genetic parameters were averaged for a standard crop of eucalypts (Table 6.1; M. Battaglia personal communication), from the known responses of three species: *Eucalyptus globulus, E. nitens* and *E. pauciflora* (Battaglia *et al.* 1996; Slatyer & Morrow 1977).

Table 6.1 Physiological parameters for a standard crop of eucalypts averaged from the published responses of *Eucalyptus pauciflora* (after Slatyer & Morrow 1977), *E. nitens* and *E. globulus* (after Battaglia *et al.* 1996), for calculating the light limited daily canopy gross photosynthetic productivity and net carbon gain (after Battaglia & Sands 1997, in prep.).

Genetic Parameter	Symbol	Value and units
The strength of the temperature dependence of the quantum efficiency, α	α_1	$0.016°C^{-1}$
Canopy extinction coefficient	k	0.5 m^2 ground m^{-2} leaf
The strength with which the optimum temperature for photosynthesis (T_{opt}) tracks the long-term mean maximum temperature ($\overline{T_x}$)	τ	0.3
The shape of the light-response curve	θ	0.95
The absolute maximum light saturated photosynthetic rate at an optimum acclimation temperature	A^*_{opt}	17 mmol m^{-2} sec^{-1}
The optimum acclimation temperature	T^*_{opt}	20°C
The acclimation temperature at which the optimum temperature and the acclimation temperature are identical	T_{pref}	23°C
The change in acclimation temperature from T^*_{opt} required to reduce the optimum photosynthetic rate by half its maximum level	$t^*_{1/2}$	12°C
The temperature change from T_{opt} required to reduce the light saturated photosynthetic rate to half the optimum value	$t_{1/2}$	12°C
Construction respiration coefficient	r_c	0.25
Value of the daily woody tissue maintenance respiration coefficient at standardised reference temperature, $T_0 = 20°C$	r_{w0}	0.35
Temperature rate constant for the dark respiration rate at long-term daily mean temperatures	k_{dav}	$0.04°C^{-1}$
Value of dark respiration at standardised reference temperature, T_0 = 20°C	r_{d0}	1 $\mu molCO_2$ m^{-2} leaf s^{-1}
Dark respiration temperature rate constant at standardised reference temperature, $T_0 = 20°C$	k_{d0}	$0.07°C^{-1}$
Temperature sensitivity of dark respiration rate constant for 10-day running mean daily temperature	k_{d1}	$0.025°C^{-1}$
Temperature constant for the daily woody tissue maintenance respiration coefficient	T_w	1.0°C

Light-saturated photosynthetic rates depend on the plant response to diurnal temperature variation (after Battaglia & Sands 1997). These responses were genetically controlled by the acclimation rate of photosynthesis to seasonally varying temperatures (τ), the dimensions of the maximal and acclimated parabolas ($t^*_{1/2}$, $t_{1/2}$), the preferred temperature (T_{pref}) at which the optimum temperature for photosynthesis and the environmental temperature correspond, and the optimal level of net photosynthesis (A^*_{opt}). Since the relative differences between sites, rather than actual estimates were of interest in gradient analyses, a value of unity for A^*_{opt} normalises the rate of light saturated photosynthesis. Typical examples of the diurnal trace of photosynthesis (A_X) based on these parameters are indicated in Figure 6.2.

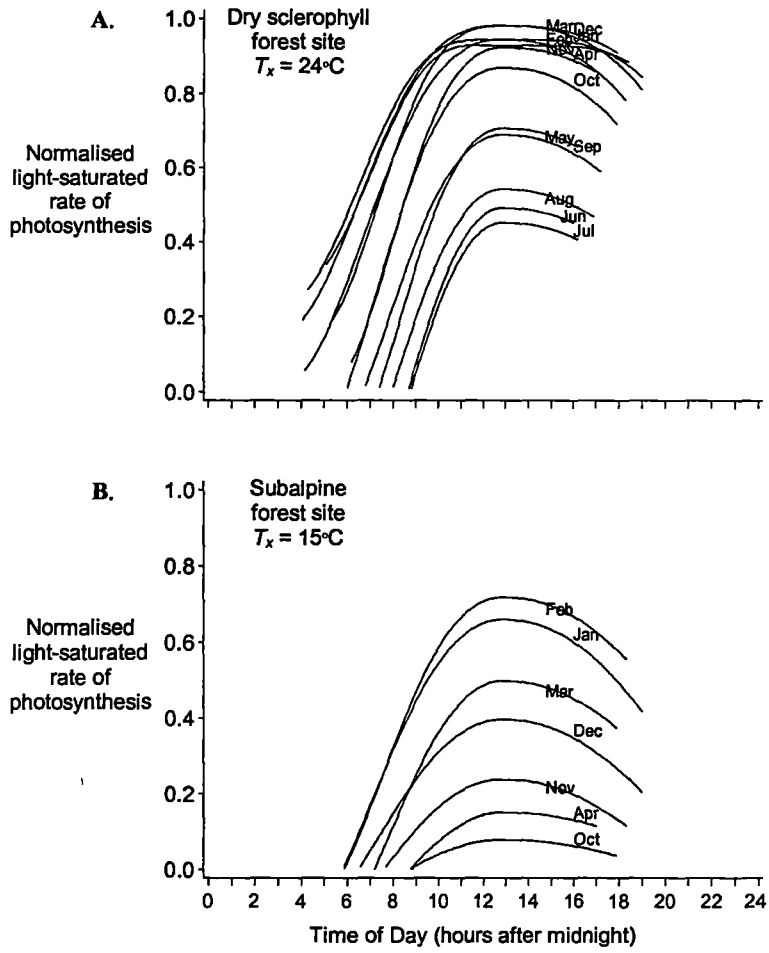


Figure 6.2A, B. Examples of seasonal variation in the diurnal trace of the light-saturated rate of photosynthesis assuming a standard crop of eucalypts (see Table 6.1).

Note that no photosynthesis occurs between May and September and the rate of photosynthesis was suppressed all year round at the subalpine forest site.

Light-limited gross canopy photosynthetic productivity combines the previously calculated temperature-dependent estimate for the light-saturated rate of photosynthesis with the diurnal variation in light. This includes the additional genetic parameters for the shape of the light response curve (θ) and the temperature sensitivity (α_1) of the quantum efficiency (α_0). The vegetation parameter for leaf area index was assumed to be large enough to intercept all available light, providing a simple bioassay of light-limited canopy gross photosynthetic site productivity (M. Battaglia personal communication).

The acquisition of carbon comes at a respiratory cost. Net canopy carbon gain is the residual of carbon uptake and respiratory losses. It is this net canopy carbon that is the substance for future plant growth and the replacement of lost tissues. It is usual to consider two types of respiration: that respiration involved in the production of dry matter (construction respiration) and that involved in maintaining dry matter (maintenance respiration). The first is often taken as a fixed proportion of carbon gain, usually 0.25 (after Penning de Vries 1974). Maintenance respiration is temperature dependent and acclimates, and is treated in the manner of Battaglia & Sands (1997). The net effect of this simple respiration model is that respiration costs increase exponentially with site temperatures. However, since photosynthesis is also increasing with mean daily site temperature over the range of sites in Tasmania, carbon use efficiency (net carbon gain/total carbon uptake) remains relatively constant (*cf.* Ryan *et al.* 1996).

6.2.4 Modelling performance of individual *Eucalyptus* species

Logistic regression models of the binomial response of *Eucalyptus regnans, E. obliqua, E. amygdalina* and *E. globulus* were used to assess the relative gains in modelling plant distributions based on direct environmental gradients or productivity/stress indices. The sampling domains for the set of presence and absence data for each species were determined by their geographic (10×10 km grid-cell scale) and altitude (50 m resolution) ranges (defined in Chapter 3).

Summary statistics (annual mean or total, maximum and minimum) were calculated from the monthly estimates for each environmental or physiological gradient. Each summary statistic represents a univariate gradient. The shape of a species' response was assessed by considering up to the fourth order polynomial. This allowed for a range of possible response shapes varying from linear to skewed or bimodal curves. The generalised linear model building strategy was based on logistic regression with backward elimination (PROC LOGISTIC, SAS Institute Inc. 1990d, 1997). The highest order polynomial for the variable with the poorest fit to the model (significance level of 0.01%, $p < 0.0001$) was removed and the model refitted. This backward elimination method was repeated until the remaining highest order polynomial was significant. Lack of significance among the lower order polynomials was ignored in favour of the significance represented by the highest order polynomial following the marginality rules of Nelder (1994).

The potential gains in the interpretability of results and the overall explanatory power for subsequent predictive modelling of species' distributions were assessed. Comparisons were made between univariate gradients in canopy productivity, thermal sum indices, and the original or derived climatic estimates for temperature and light. Univariate models were compared by their overall deviance and degrees of freedom, representing the complexity of shape for the ecological response.

Observed and predicted responses to selected univariate gradients for temperature or productivity were graphically displayed. The observed responses were approximations of the actual binomial response, calculated from the frequency of species' presence relative to the presence and absence in 10 evenly spaced classes of the environmental gradient, and displayed by their class means. The observed responses provide an indication of the 'residual difference' between actual observations and predicted values. Bivariate models for two indicative temperature variables demonstrated the patterns of environmental response among the four *Eucalyptus* species.

Since the relative significance of gradients in temperature, light or photosynthetic productivity differ depending on the presence of other factors for water and nutrients, these variables were tested in multivariate models. Multivariate models were derived by backward elimination of potential maximal models comprising mean annual variables (all with initially up to their fourth order polynomials) from all major environmental factors — water, temperature, nutrients, and

light — tested in the presence of a covariate for sample size (plot area, ha) since this will influence the probability of occurrence. Each group of variables representing either productivity gradients, thermal sum indices or climatic factors for light and temperature were tested with a nutrient index derived from substrate categories (after Nix *et al.* 1992), and climatic water variables (rainfall, evaporation and rain days). These multivariate models were compared by a number of different model-fit statistics: overall deviance and degrees of freedom; the rank correlation between observed responses and predicted probabilities; a generalised estimate for R^2; the area under the receiver-operator characteristic curve; and a classification table that summarises model sensitivity, specificity, and rates of false negative or false positive predictions (after SAS Institute Inc. 1990d, 1997).

6.3 Results

6.3.1 Environmental gradients in temperature

The variation in direct and derived temperature gradients in a sample of Tasmanian eucalypt forest habitats are indicated in Figures 6.3 to 6.5. The potential acclimation regimes appear to offer a novel approach to the description of environmental temperatures (Fig. 6.3C). Other indices follow the generally expected patterns of minimum in winter and maxima in summer (Figs. 6.3A, 6.3B, 6.4, 6.5). However, the rate variables for temperature follow a different pattern that also seems to vary significantly between equinoctial seasons (autumn and spring). These rate regimes also demonstrate some of the differences between the day-time maximum and day-time minimum changes in temperature during the solstice seasons of winter and summer, as well as autumn and winter. For example, the autumn night-time acclimation environment changes faster between April/May (or March/April in some habitats), but in spring this is delayed until October/November. The day-time temperature maximum spring fastest rate of change is earlier (September/October) than the night-time equivalent. These figures suggest that there may be novel information about environmental temperature regimes from the rates of change variables that is not otherwise included in the more conventional temperature estimates.

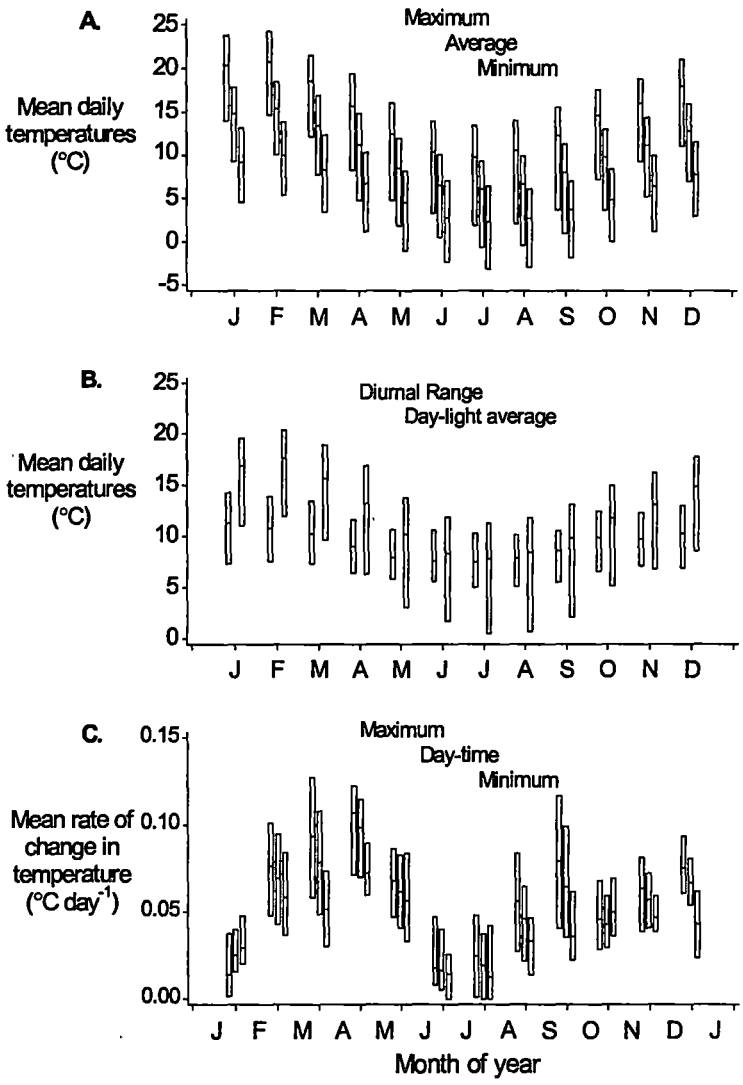


Figure 6.3. Seasonality and monthly range of mean daily temperatures for forest habitats dominated by *Eucalyptus* species in Tasmania (n = 15640). **A.** Maximum, average, minimum. **B.** Diurnal range and daylight averages (mean of temperature variation during daylight hours). **C.** Rate of seasonal temperature changes (°C day^{-1}) in daily maximum or minimum, or average daytime temperatures. Estimates of average daily maximum and minimum temperature per month were obtained from ESOCLIM (McMahon *et al.* 1996). Bars indicate maximum, minimum and mean values. Diurnal temperature variation was calculated from the sine curve fits to quasi-daily values of maximum and minimum temperature (see Box 6.1) using the monthly daylight hours (middle day of the month) obtained from an excel worksheet adapted from Iqbal (1983) by M. Fowler (personal communication, October 1995).

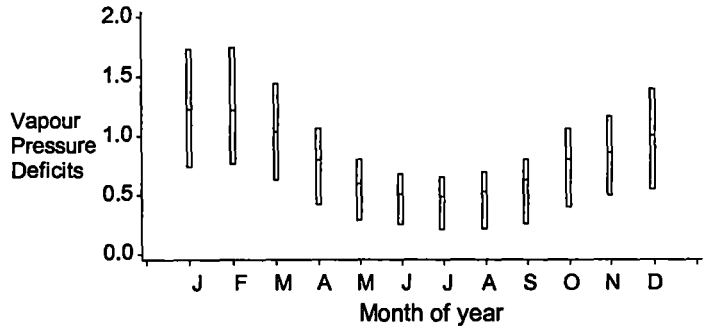


Figure 6.4. Seasonality and monthly range of vapour pressure deficit (KPa, after Landsberg 1986, see Box 6.2) for a sample of forest habitats dominated by *Eucalyptus* species in Tasmania (n = 15640). Bars indicate maximum, minimum and mean values.

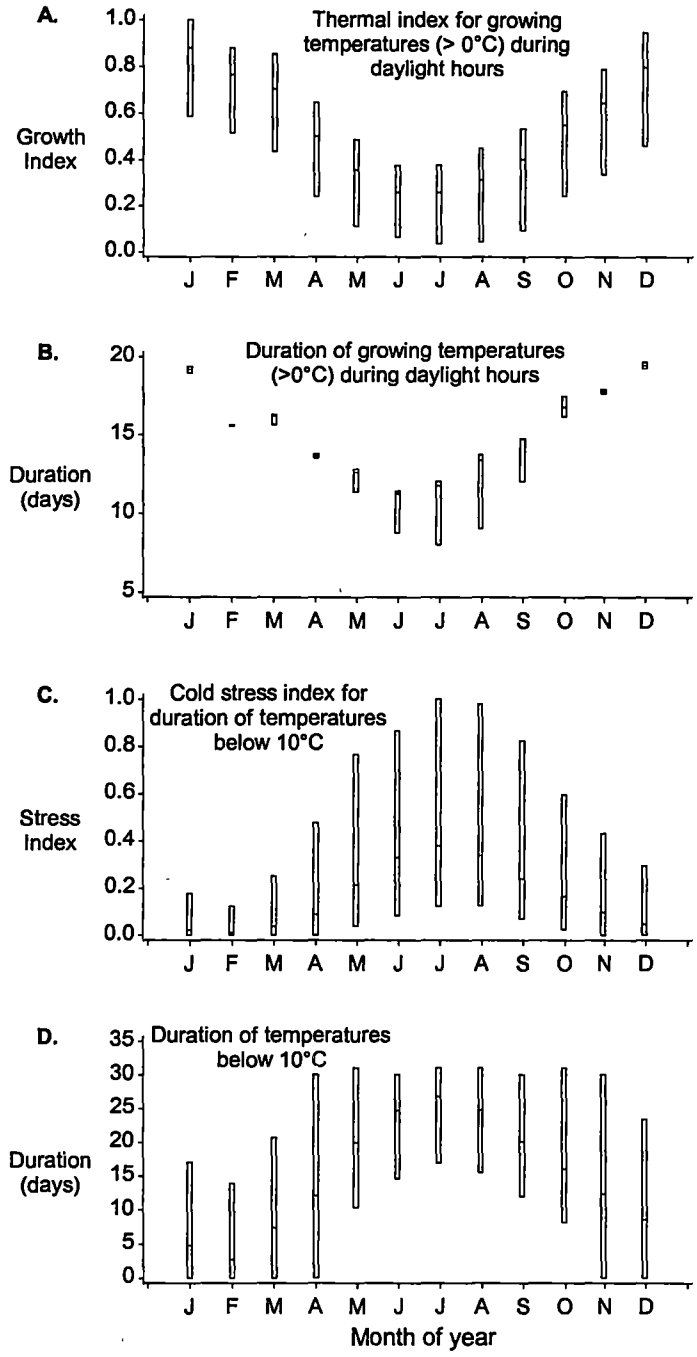


Figure 6.5 Seasonality and monthly range of the thermal indices for forest habitats dominated by *Eucalyptus* species in Tasmania (n = 15640).

A & B. Duration and magnitude of growing temperatures above 0°C during daylight hours (see Box 6.2).

C & D. Duration and intensity of cold stress below 10°C during day and night hours.

Note that the number of days per month varies between 28 and 31. The accumulated degree days in each case were normalised (0-1) to the range. Bars indicate maximum, minimum and mean values.

Estimates for monthly daylight hours (middle day of the month) were obtained from an Excel worksheet adapted from Iqbal (1983), by M. Fowler (personal communication, October 1995).

6.3.2 Eucalypt canopy productivity

The seasonal trace of the optimal rate of normalised light-saturated photosynthesis (A_{opt}) and the acclimation temperature (T_{opt}), with the mean maximum daily temperature per month and the relative rate of photosynthesis (based on the average day-light temperature), are demonstrated for all forest sites (Fig. 6.6).

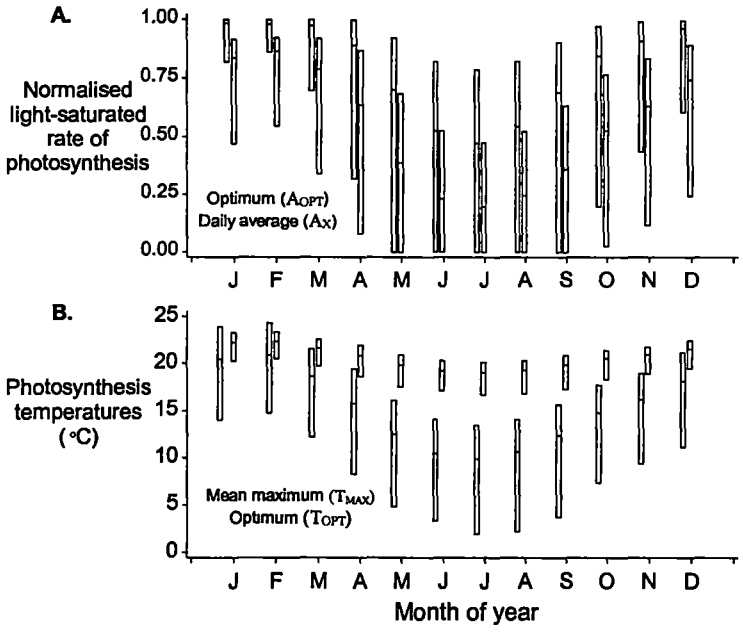


Figure 6.6 Seasonality and monthly range of the average daily rate of light-saturated photosynthesis for a standard crop of eucalypts (see Table 6.1), estimated for a sample of forest habitats dominated by *Eucalyptus* species in Tasmania (n = 15640). **A.** Normalised values for light-saturated photosynthesis (A_x) and optimum photosynthesis (A_{opt}). **B.** Mean maximum temperatures (T_X) and photosynthetic optimum temperature (T_{opt}). Bars indicate maximum, minimum and mean values.

The differences in the estimation of gross canopy productivity when temperature is constant, and only light is varying, was compared with light-limited canopy productivity (Fig. 6.7a). The normalised values for light-limited canopy productivity are indicated in Figure 6.7b. The seasonal variation in leaf respiration and canopy carbon gain shows that respiration is acclimating and that carbon use efficiency is relatively constant resulting in the dominance of photosynthesis among these canopy processes (Fig. 6.7b).

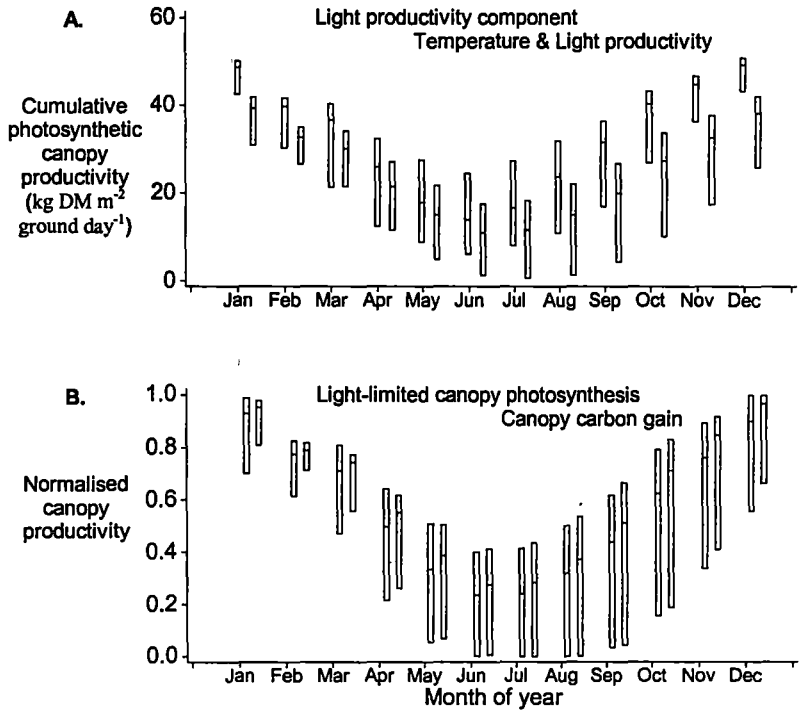


Figure 6.7 Seasonality and monthly range of an indices for canopy productivity in forest habitats dominated by *Eucalyptus* species in Tasmania (n = 15640). **A.** Comparison of indices for non-temperature limited, and temperature and light-limited, photosynthetic productivity. **B.** Normalised index for light-limited canopy photosynthetic productivity and canopy carbon gain. Relative rate of photosynthesis or respiration were calculated by assuming a standard crop of eucalypts (see Table 6.1). Bars indicate maximum, minimum and mean values.

6.3.3 Correlation between environmental variables and productivity indices

The relationship between annual statistics for temperature and light, and derived indices for thermal sum and physiological processes are demonstrated using a Pearson correlation matrix (Table 6.2). These indices represent the average of within-year mean monthly variation for forested sites in Tasmania. The original temperature and light estimates were relatively uncorrelated, but some derived variables for temperature were highly correlated with each other. For example, the thermal sum for temperatures above 0°C was, on average, effectively the same as the mean maximum temperature, and the light-saturated rate of photosynthesis was effectively the same as the average day-time temperature. This close correspondence to some extent reflects the numerical dominance of day-time (mean maximum) temperature regimes in the calculation of mean annual indices.

Table 6.2 Pearson correlation coefficients (percentages) for mean annual statistics of the monthly thermal variables for a sample of eucalypt forest habitat in Tasmania (n = 15 640). Radiation (Rq), maximum temperature (Tmax), minimum temperature (Tmin), temperature range (Trange), day-time average temperature (Tday), rate of change in maximum temperature (Tμmax), rate of change in minimum temperature (Tμmin), vapour pressure deficit (VPD), thermal growth sum (GDD0), thermal cold-stress sum (SDD10), light-saturated rate of photosynthesis (Ax), light-limited canopy gross photosynthetic productivity (Cpn), canopy carbon gain (Gc), rainfall (PT), evaporation (EP), rain-days (DP). The set of variables which define mean annual conditions were indicated by '‡' and the set of variables which approximate seasonal conditions were indicated by '†'.

	R_q	T_{max}	T_{min}	T_{range}	T_{day}	$T\mu_{max}$	$T\mu_{min}$	V_{PD}	GDD_0	SDD_{10}	A_x	C_{pn}	G_c	P_T	E_P	D_P
R_q	*100*	19	3	37	16	43	73	33	18	-2	10	20	19	-23	61	-77
T_{max}‡		*100*	89	56	99	-31	55	89	100	-93	98	98	97	-35	78	-47
T_{min}‡			*100*	12	94	-67	23	59	93	-97	96	93	90	-26	72	-20
T_{range}†				*100*	45	53	77	87	48	-26	40	45	46	-28	39	-67
T_{day}‡					*100*	-41	48	83	100	-96	100	99	97	-33	78	-41
$T\mu_{max}$†						*100*	44	11	-39	58	-48	-41	-39	9	-11	-29
$T\mu_{min}$†							*100*	77	50	-24	40	45	41	-40	64	-80
V_{PD}†								*100*	85	-68	79	82	81	-36	68	-64
GDD_0‡									*100*	-95	99	99	97	-34	78	-44
SDD_{10}‡										*100*	-98	-97	-97	24	-71	26
A_x‡											*100*	99	98	-30	75	-36
C_{pn}‡												*100*	99	-31	79	-45
G_c													*100*	-28	77	-45
P_T‡														*100*	-39	60
E_P‡															*100*	-61
D_P‡																*100*

Other temperature variables were relatively unrelated, especially the mean temperature range and mean minimum temperature (correlation = 12%). Overall the temperature and productivity indices were poorly correlated with rainfall, but moderately correlated with evaporation. Most temperature variables were also poorly correlated with mean annual solar radiation, particularly the minimum temperature. These correlations indicate that several gradients may be needed to define the environmental regime of temperature from mean annual and seasonal changes in conditions (e.g. Fig. 6.8). However, they do not necessarily indicate which factors would be most important in explaining the binary response of species' distributions in compiled ecological data.

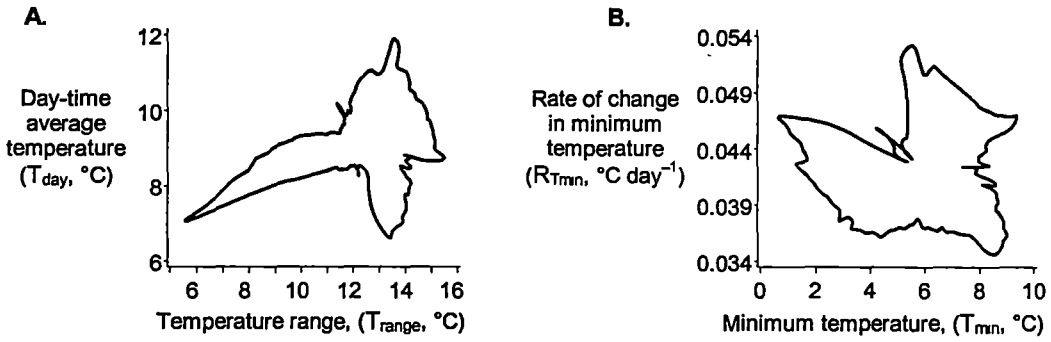


Figure 6.8 Examples of the two dimensions of temperature derived from annual and seasonal sets of conditions. Data envelopes for the sample of eucalypt forest occurrences in Tasmania (n = 15 640). **A.** Mean annual day-time average temperature versus mean annual temperature range. **B.** Mean annual rate of change in minimum temperature versus mean annual minimum temperature .

6.3.4 Modelling performance of individual *Eucalyptus* species

The change in deviance in a logistic regression for the fit of each set of statistics (annual mean, or monthly maximum or minimum) for temperature, light, and derived indices for thermal sum and physiological productivity were summarised for four *Eucalyptus* species (Table 6.3). Some of these responses are illustrated in Figure 6.9.

Most species have a highly significant univariate response to an index for day-time temperature and the summation of growing degree-days for day-time temperature (mean, minimum or maximum). However, their response to these factors depends upon whether it reflects an annual mean or absolute minimum or maximum. This individualistic response reflects some of the higher changes in deviance recorded for other variables. For example, *E. amygdalina* was strongly associated with the maximum rate of change in minimum temperature, consistent with its response to absolute maximum temperature and the maximum day-time temperature. This response to the rate of change in minimum temperature reflects the reversal in correlation between these variables (compare Pearson coefficients for T_{max} and R_{Tmin} in Table 6.2). The rate of change in temperature may assist in delineating its distribution between inland and coastal sites where seasonal factors influencing the degree of drought are a component of its habitat (e.g. Duncan & Brown 1985). Rate variables therefore appear promising as a temperature index for correlating some species' distribution patterns, whatever the causal relationship which underlies its significance.

Table 6.3 Change in deviance (logistic regression) for variables defining thermal conditions of the environment or a physiologically defined equivalent for observations of the presence (n_1) and absence (n_0) of four *Eucalyptus* species.

The degrees of freedom (df) indicate the number of polynomial elements (up to the fourth order) that define the shape of the response to each variable. Results were only shown for polynomial fits for each variable that was significant at the 0.01% ($p < 0.001$) level or greater (else fits were not significant, *ns*). The most significant responses are indicated in bold. The thermal variables that describe mean annual conditions are indicated by '‡' and the thermal variables for seasonal conditions are indicated by '†'.

Variable	*E. regnans* (n_1 = 2462) (n_0 = 6713)		*E. obliqua* (n_1 = 8182) (n_0 = 5433)		*E. amygdalina* (n_1 = 3986) (n_0 = 8321)		*E. globulus* (n_1 = 1091) (n_0 = 3980)	
	Δ *Dev* (NULL=10 672)	df	Δ *Dev* (NULL=18 316)	df	Δ *Dev* (NULL=15 501)	df	Δ *Dev* (NULL=5280)	df
Environmental temperature, solar radiation and day length variables:								
Flat surface solar radiation (R_q, W m^{-2})								
Mean annual	712	4	1334	4	2115	3	94	4
Maximum month	239	3	1160	4	2134	2	116	3
Minimum month	**778**	4	1196	4	1935	4	109	4
Maximum temperature‡ (T_{max}, °C)								
Mean annual	369	2	2323	2	2236	3	171	3
Maximum month	256	4	2293	4	**2576**	4	141	2
Minimum month	**705**	4	**2375**	4	1274	3	185	3
Minimum temperature‡ (T_{min}, °C)								
Mean annual	667	4	2134	3	707	4	**210**	4
Maximum month	594	4	2039	3	1469	4	190	4
Minimum month	588	3	**2281**	3	450	4	192	3
Average daylight temperature (T_{day}, °C)								
Mean annual	488	2	**2358**	2	1862	3	176	3
Maximum month	282	2	2283	4	**2604**	3	159	3
Minimum month	**755**	4	2221	2	865	3	**203**	4
Diurnal range between maximum and minimum temperatures† (T_{range}, °C)								
Mean annual	337	2	329	3	1759	4	107	2
Maximum month	272	2	873	4	431	2	43	3
Minimum month	267	2	422	4	2206	4	126	4
Rate of change in maximum temperature† ($T\mu_{max}$, °C day^{-1})								
Mean annual	319	2	892	4	447	1	29	1
Maximum month	268	3	795	3	354	2	54	3
Minimum month	319	4	1315	3	1562	4	35	3
Rate of change in minimum temperature† ($T\mu_{min}$, °C day^{-1})								
Mean annual	104	3	400	3	2606	4	80	4
Maximum month	146	3	208	4	**2912**	4	44	2
Minimum month	328	4	116	2	1164	3	48	1
Vapour pressure deficit† (V_{PD})								
Mean annual	232	2	1149	4	2273	4	159	4
Maximum month	329	2	395	4	2081	4	171	2
Minimum month	364	2	2081	3	2297	3	152	3
Thermal sum variables:								
Growth-index for daylight hours above 0°C‡ (GDD_0, *day °C*)								
Mean annual	438	2	**2345**	2	1962	3	172	3
Maximum month	235	2	2263	4	**2408**	3	151	3
Minimum month	**721**	4	2177	3	996	3	**198**	4
Cold stress-index for day and night hours below 10°C‡ (SDD_{10}, day °C)								
Mean annual	664	4	2194	4	664	3	**199**	3
Maximum month	**684**	2	**2288**	3	601	4	196	3
Minimum month	314	4	2034	2	815	4	163	2
Canopy photosynthesis and productivity variables:								
Light-saturated rate of photosynthesis‡ (*i.e. temperature-limited*, A_x)								
Mean annual	588	4	**2389**	4	1360	4	186	3
Maximum month	396	4	2111	4	1213	3	179	3
Minimum month	**761**	4	2249	4	940	3	200	4
light-limited canopy gross photosynthetic productivity‡ (*i.e. temperature and light-limited*, C_{pn})								
Mean annual	369	2	2214	3	1755	4	175	3
Maximum month	302	4	2020	3	1874	4	165	3
Minimum month	512	4	2108	2	1480	3	193	4
Carbon gain‡ (*leaf respiration minus light-limited canopy gross photosynthetic productivity*, G_c)								
Mean annual	282	2	2203	4	1627	4	179	3
Maximum month	216	3	1514	4	1159	3	130	3
Minimum month	434	4	2138	2	1661	3	188	4

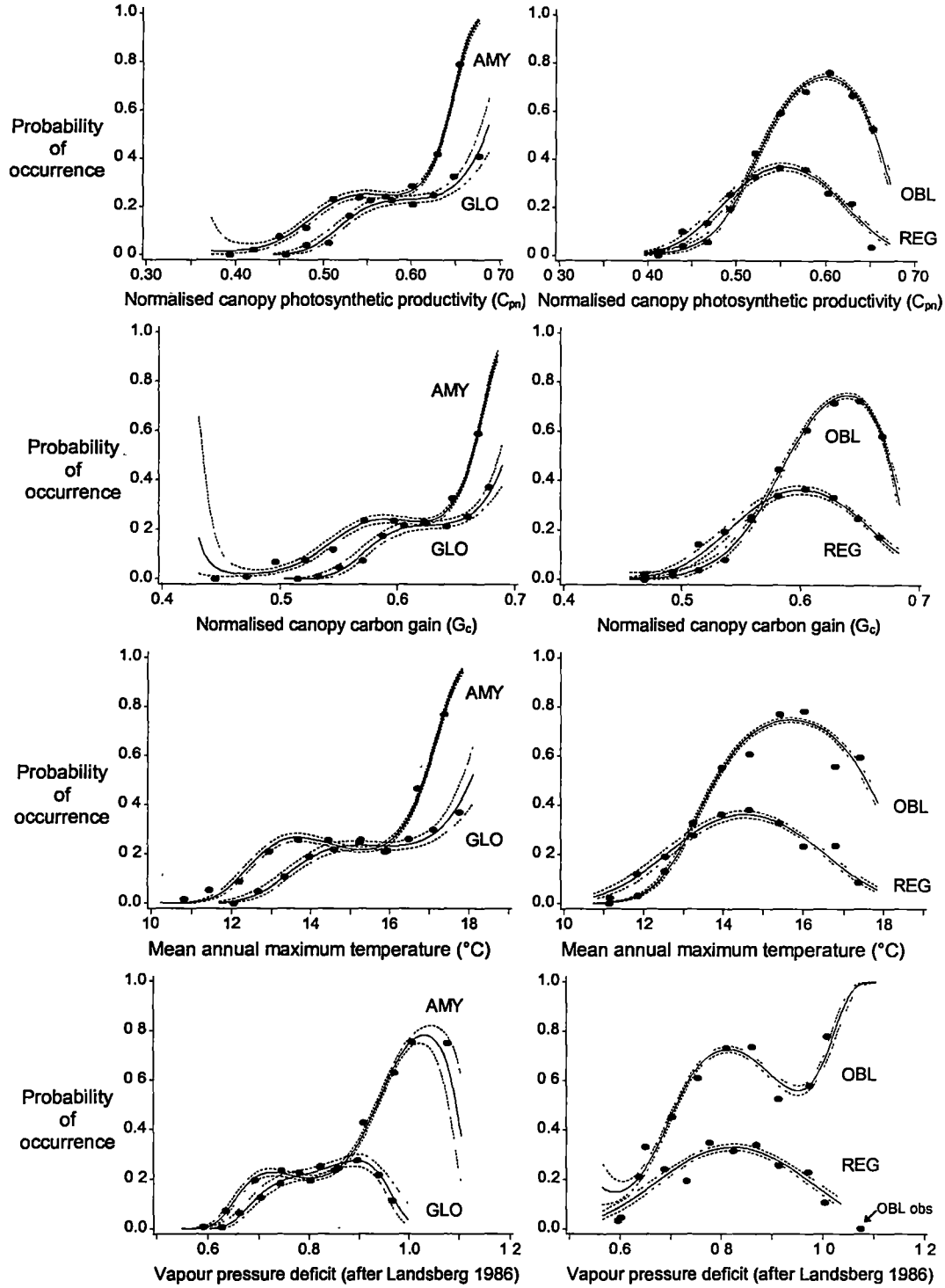


Figure 6.9. Univariate responses to indicative temperature and productivity variables (after Table 6.3). Comparison of predicted responses for four Eucalyptus species: AMY (E. amygdalina), GLO (E. globulus), OBL (E. obliqua), REG (E. regnans). Mean annual indices for: light-limited canopy photosynthetic productivity (Cpn), canopy carbon gain (Gc), maximum temperature (Tmax), and vapour pressure deficit (VPD). Observations (filled dot) were shown for the frequency of presences relative to presences and absences in 10 equal classes for each variable (positioned by class means).

In each case, a species' correlative response to the estimate for vapour pressure deficit (from Landsberg 1986) was an improvement over the diurnal temperature range. The univariate productivity gradients derived from the physiological responses for photosynthesis improved the explanation of species' distributions in only two cases (Table 6.3). There were slight

improvements in the responses for *E. regnans* and *E. obliqua*. However, the responses for *E. globulus* and *E. amygdalina* were poorer than the original temperature variables. This is opposite to the expected response given the origin of the parameters largely from the physiology of *E. globulus* (i.e. Battaglia *et al.* 1996; Battaglia & Sands 1997). These differences in model fits between species suggests that the genetic parameters for a standard crop of eucalypts may not be equally applicable to all species when comparing responses along a generalised productivity gradient. Where there are gains in univariate explanation achieved by using the physiological derivatives for productivity, these have been small and may be little more than a statistical coincidence. A multivariate approach to analysis is needed to confirm these trends.

The different types of response of species to two relatively uncorrelated temperature variables, mean annual average day-time temperature and mean annual temperature range, are indicated in Table 6.4 (displayed in Fig. 6.10). For consistency of comparison, the variable preferences indicated in Table 6.3 were ignored (but it should be noted that average day-time temperature in particular was a good predictor for all species). Because the sampling domain for each species was different (after Williams & Potts 1996), the shape of the envelope for the two temperature variables also differs (compare with complete data range in Figure 6.8). The generalised R^2 ($\widetilde{R}^2$) suggests that these temperature variables were most significant to the distribution of the widespread species, *E. amygdalina* and *E. obliqua*, and least significant to the less widespread distributions of *E globulus* and *E. regnans* (Table 6.4). For each species, the bivariate models for temperature were not adequate fits to the data ($p < 0.01$ for the Hosmer and Lemeshow goodness-of-fit test). Not surprisingly, more than two variables are needed to adequately define the fit of these models. Nevertheless, the graphic display of the ecological responses to these two temperature gradients provides opportunities for interpreting the distribution patterns in terms of a comparison of habitats for these species. A more complete model of the environment associated with the distribution of each species is needed to clarify their responses, and this is carried out below.

Table 6.4 Bivariate responses to temperature. The response of each species is compared for mean annual average day-time temperature (T_{day}) and mean annual temperature range (T_{range}).
The NULL model statistics and number of observation (presences and absences) were given in Table 6.3. The linear predictor (η) is given for the fit of the logistic regression functions. Response shapes for each variable were defined by the order of the polynomial. Results were only shown for fits that were significant at the 0.01% ($p < 0.001$) level or greater, following backward selection of the maximum number of variables. Results were only shown for fits that were significant at the 0.01% ($p < 0.001$) level or greater, following backward selection of the maximum number of variables. The model fits are indicated by the *Degrees of Freedom* (df), *Change in Deviance* (ΔDev), the *Association of Predicted Probabilities and Observed Responses* (concordant, discordant and tied responses), the *Classification Table for Trials of Event/Noevent* (based upon the response being an event if the probability of occurrence is greater than 0.5, indicating the percentage correct, and rates of false positive, F+, and false negative, F–, predictions), the area under the receiver-operator characteristic curve (C), the Hosmer and Lemeshow goodness-of-fit test [χ^2 for 8 df; Hosmer & Lemeshow (1989) with $p < 0.01$ denoted by '*' suggesting that the fitted model is not an adequate model, otherwise the fit is adequate], and the generalised, maximum rescaled R^2 measure for the fitted model (denoted $\tilde{R}^2$). Detail of the logistic regression method is given in SAS Institute Inc. (1990d).

Model	df	ΔDev	score	conc	disc	tied	%correct	F+	F–	c	% $\tilde{R}^2$	χ^2
Eucalyptus regnans:												
$\eta = -3242.7 + 238.0\times T_{day} - 29.7159\times T_{day}^2 + 1.6509\times T_{day}^3 - 0.0345\times T_{day}^4$												
$+ 1107.4\times T_{range} - 181.8\times T_{range}^2 + 13.2554\times T_{range}^3 - 0.3617\times T_{range}^4$												
	8	710	595	65.7	33.6	0.6	73.2	.	26.8	0.661	10.9	101*
Eucalyptus obliqua:												
$\eta = -898.1 + 10.1818\times T_{day} - 0.3915\times T_{day}^2$												
$+ 368.4\times T_{range} - 60.1934\times T_{range}^2 + 4.3167\times T_{range}^3 - 0.1148\times T_{range}^4$												
	6	2711	2370	72.9	26.6	0.5	70.1	30.6	27.7	0.732	24.5	108*
Eucalyptus amygdalina:												
$\eta = -1867.3 - 112.0\times T_{day} + 16.9727\times T_{day}^2 - 1.1089\times T_{day}^3 + 0.0265\times T_{day}^4$												
$+ 871.5\times T_{range} - 133.4\times T_{range}^2 + 9.0617\times T_{range}^3 - 0.2301\times T_{range}^4$												
	8	2761	2631	76.6	23.1	0.3	75.6	30.2	22.9	0.767	28.1	185*
Eucalyptus globulus:												
$\eta = -224.0 + 39.3163\times T_{day} - 3.0327\times T_{day}^2 + 0.0780\times T_{day}^3$												
$+ 11.4716\times T_{range} - 0.6216\times T_{range}^2$												
	5	200	156	60.5	38.0	1.5	78.3	81.3	21.5	0.612	6.0	29*

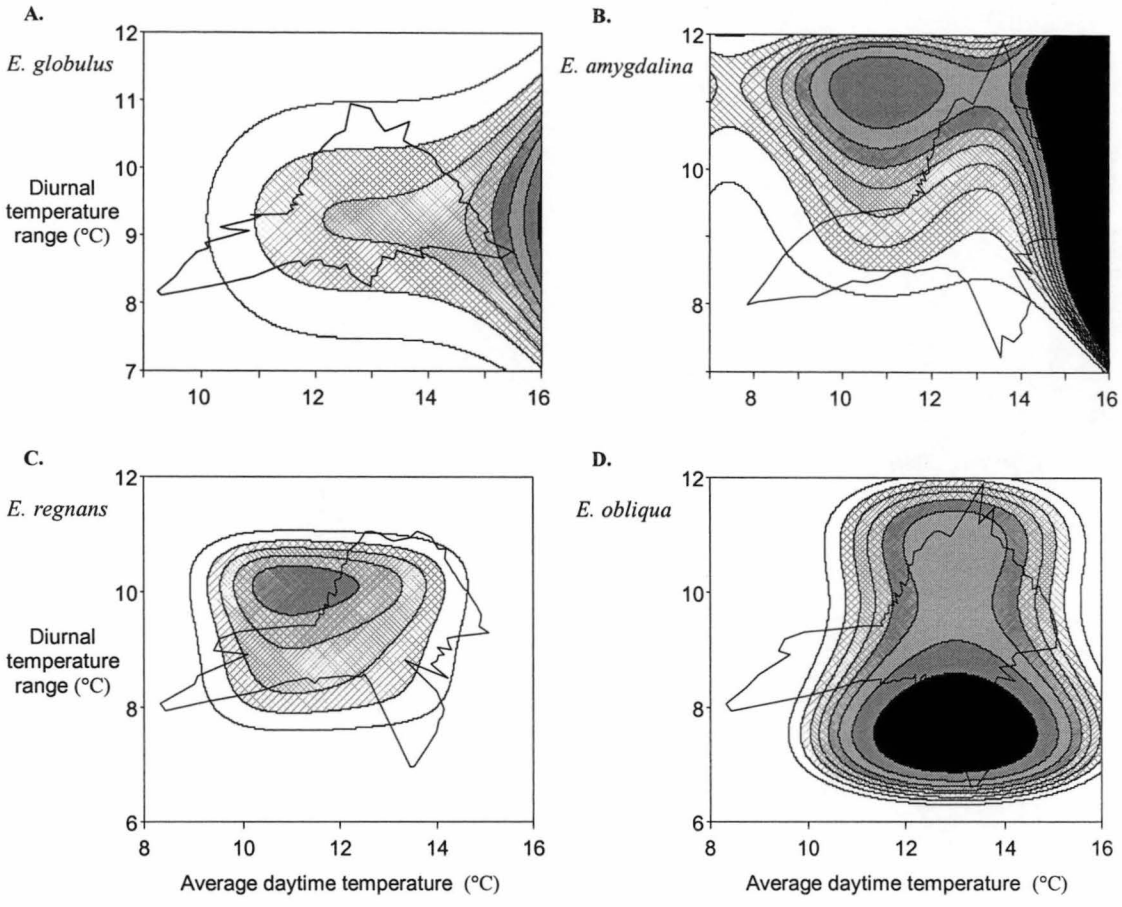


Figure 6.10 Bivariate responses to temperature. The response of each species is compared for mean annual average day-time temperature (T_{day}) and mean annual temperature range (T_{range}). Probability contours were given in 0.1 intervals starting between the first blank contour and upto 1.0 for the black pattern (from models in Table 6.4). Note the different scales. The envelope represents the data domain for each species, based on the respective geographic and environmental ranges (see Chapter 3).

The relative importance of the different temperature and productivity indices in the presence of other environmental gradients was demonstrated by the multivariate responses for the four *Eucalyptus* species (Table 6.5). In each case, complex models were needed to describe species ecological responses. Variables for annual and seasonal factors, and seemingly highly correlated variables were significant together. For example, all canopy productivity indices were always significant, although usually with opposing influences between photosynthesis and carbon gain variables.

Table 6.5 Multivariate responses. The response of each species to direct environmental gradients in temperature [mean annual minimum (T_{min}) and maximum temperature (T_{max}), mean annual temperature range (T_{range}), rate of change in minimum (R_{Tmin}) or maximum temperatures (R_{Tmax})] or light (R_q) is compared with the normalised productivity gradients based on photosynthesis [light-saturated rate of photosynthesis (A_x), light-limited canopy photosynthetic productivity (C_{pn}), or canopy carbon gain between photosynthesis and respiration (not normalised G_c)] and vapour pressure deficit (V_{PD}), or the normalised thermal sum indices [growth index (GDD_0), cold stress index (SDD_{10})], all in the context of other environmental gradients for water [rainfall (P_T), evaporation (E_P), rain days (D_P)] and nutrients (substrate nutrient index, N). Sample size (S_A, ha) was included as a covariate in each case.

Set 1: $S_A + T_{min} + T_{max} + T_{range} + R_{Tmin} + R_{Tmax} + R_q + P_T + E_P + D_P + N$ — Environmental gradients.
Set 2: $A_x + C_{pn} + G_c + V_{PD} + P_T + E_P + D_P + N$ — Environmental gradients and photosynthetic productivity.
set 3: $GDD_0 + SDD_{10} + V_{PD} + R_q + P_T + E_P + D_P + N$ — Environmental gradients and thermal sum indices.

The NULL model statistics and number of observation (presence and absence) were given in Table 6.3. The linear predictor (η) is given for the fit of the logistic regression functions. Response shapes for each variable were defined by the order of the polynomial. Results were only shown for fits that were significant at the 0.01% ($p < 0.001$) level or greater, following backward selection of the maximum number of variables. The model fits are indicated by the *Degrees of Freedom* (df), *Change in Deviance* (Δ *Dev*), the *Association of Predicted Probabilities and Observed Responses* (concordant, discordant and tied responses), the *Classification Table for Trials of Event/Noevent* (based upon the response being an event if the probability of occurrence is greater than 0.5, indicating the percentage correct, and rates of false positive, F+, and false negative, F–, predictions), the area under the receiver-operator characteristic curve (C), the Hosmer and Lemeshow goodness-of-fit test [χ^2 for 8 df; Hosmer & Lemeshow (1989) with $p < 0.01$ denoted by '*' suggesting that the fitted model is not an adequate model, otherwise the fit is adequate], and the generalised, maximum rescaled R^2 measure for the fitted model (denoted $\tilde{R}^2$). Detail of the logistic regression method is given in SAS Institute Inc. (1990d).

Eucalyptus regnans:

$$\eta = -6783.37 + 148.924\times S_A - 894.304\times S_A^2 + 2012.77\times S_A^3 - 1452.92\times S_A^4 + 0.0\times T_{min} - 20.9648\times T_{min}^2 + 2.16198\times T_{min}^3 - 0.082437\times T_{min}^4 + 727.881\times T_{max} - 69.6603\times T_{max}^2 + 3.35882\times T_{max}^3 - 0.060287\times T_{max}^4 - 76.3065\times T_{range} - 1.00047\times T_{range}^2 + 591.068\times R_{Tmax} + 52697.44\times R_{Tmin} - 1183.2886\times 10^3\times R_{Tmin}^2 + 8931.8034\times 10^3\times R_{Tmin}^3 + 790.841\times R_q - 57.6851\times R_q^2 + 1.39813\times R_q^3 + 1.2389\times 10^{-2}\times P_T - 4.7192\times 10^{-6}\times P_T^2 - 2.2059\times 10^{-2}\times E_P - 2.1216\times 10^{-2}\times D_P + 18.5950\times N - 71.9964\times N^2 + 98.4966\times N^3 - 43.8414\times N^4$$

Model	df	Δ *Dev*	score	conc	disc	tied	%correct	F+	F–	C	% $\tilde{R}^2$	χ^2
Environmental gradients:	28	2268	1880	80.6	19.2	0.2	77.6	37.2	19.3	0.807	31.9	25*

$$\eta = -3317.9 + 124.8\times S_A - 754.6\times S_A^2 + 1711.1\times S_A^3 - 1244.7\times S_A^4 + 43901.1\times A_x - 95891.2\times A_x^2 + 92481.4\times A_x^3 - 33397.6\times A_x^4 + 733.3\times C_{pn} - 1444.9\times G_c + 111.6\times G_c^2 - 2.9111\times G_c^3 + 4387.9\times V_{PD} - 7320.9\times V_{PD}^2 + 5457.3\times V_{PD}^3 - 1536.7\times V_{PD}^4 - 0.0198\times P_T + 1.9\times 10^{-5}\times P_T^2 - 5.57\times 10^{-9}\times P_T^3 + 2.7885\times E_P - 4.44\times 10^{-3}\times E_P^2 + 3.082\times 10^{-6}\times E_P^3 - 7.95\times 10^{-10}\times E_P^4 + 1.6893\times D_P - 7.79\times 10^{-3}\times D_P^2 + 1.1\times 10^{-5}\times D_P^3 + 14.0006\times N - 55.1602\times N^2 + 75.9637\times N^3 - 33.6803\times N^4$$

Model	df	Δ *Dev*	score	conc	disc	tied	%correct	F+	F–	C	% $\tilde{R}^2$	χ^2
Productivity gradients:	30	2285	1888	80.8	19.0	0.2	78.5	35.5	18.4	0.809	32.1	35*

$$\eta = -4918.6 + 150.2\times S_A - 905.7\times S_A^2 + 2046.7\times S_A^3 - 1482.1\times S_A^4 - 6089.8\times GDD_0 + 12639.7\times GDD_0^2 - 8155.3\times GDD_0^3 + 696.3\times SDD_{10} - 2205.1\times SDD_{10}^2 + 5346.0\times SDD_{10}^3 - 8095.6\times SDD_{10}^4 + 1035.7\times R_q - 74.6519\times R_q^2 + 1.7921\times R_q^3 + 1106.2\times V_{PD} - 1331.0\times V_{PD}^2 + 488.6\times V_{PD}^3 - 0.0129\times P_T + 1.4\times 10^{-5}\times P_T^2 - 4.61\times 10^{-9}\times P_T^3 + 2.3925\times E_P - 4.08\times 10^{-3}\times E_P^2 + 3.023\times 10^{-6}\times E_P^3 - 8.26\times 10^{-10}\times E_P^4 + 2.5755\times D_P - 0.0131\times D_P^2 + 2.2\times 10^{-5}\times D_P^3$$

Model	df	Δ *Dev*	score	conc	disc	tied	%correct	F+	F–	C	% $\tilde{R}^2$	χ^2
Thermal sum indices:	27	2201	1838	80.2	19.6	0.2	77.2	39.0	19.3	0.803	31.0	57*

Table 6.5 Continued

Model	df	ΔDev	score	conc	disc	tied	%correct	F+	F–	c	$\%\widetilde{R}^2$	χ^2
Eucalyptus obliqua:												
$\eta = -177.835 - 55.0281\times S_A + 566.208\times S_A^2 - 2133.56\times S_A^3 + 2624.60\times S_A^4$ $+ 0.0\times T_{mn} + 1.62200\times T_{mn}^2 - 0.097465\times T_{mn}^3$ $+ 36.4367\times T_{max} - 2.82773\times T_{max}^2 + 0.059090\times T_{max}^3$ $- 89.0431\times T_{range} + 10.2496\times T_{range}^2 - 0.35890\times T_{range}^3$ $+ 1189.0150\times 10^2\times R_{Tmax} - 2948.0855\times 10^3\times R_{Tmax}^2 + 3230.7486\times 10^4\times R_{Tmax}^3 - 1320.1258\times 10^5\times R_{Tmax}^4$ $+ 20040.75\times R_{Tmin} - 4875.6704\times 10^2\times R_{Tmin}^2 + 3903.1542\times 10^3\times R_{Tmin}^3$ $- 475.166\times R_q + 33.2659\times R_q^2 - 0.77703\times R_q^3$ $+ 6.0286\times 10^{-2}\times P_T - 6.3676\times 10^{-5}\times P_T^2 + 2.8627\times 10^{-8}\times P_T^3 - 4.57\times 10^{-12}\times P_T^4$ $+ 8.7505\times 10^{-3}\times E_P$ $+ 9.47083\times D_P - 7.5497\times 10^{-2}\times D_P^2 + 2.6590\times 10^{-4}\times D_P^3 - 3.5010\times 10^{-7}\times D_P^4$ $- 0.46084\times N$												
Environmental gradients:	32	4880	4200	82.5	17.4	0.2	74.9	23.3	28.6	0.826	40.8	13
$\eta = -1281.0 - 49.0202\times S_A + 543.0\times S_A^2 - 2135.6\times S_A^3 + 2691.3\times S_A^4$ $+ 10174.2\times A_x - 23931.5\times A_x^2 + 25748.7\times A_x^3 - 10436.5\times A_x^4$ $+ 12568.9\times C_{pn} - 25427.1\times C_{pn}^2 + 16134.2\times C_{pn}^3$ $+ - 721.9\times G_c + 62.3120\times G_c^2 - 1.7439\times G_c^3$ $+ 507.1\times V_{PD} - 609.7\times V_{PD}^2 + 242.3\times V_{PD}^3$ $+ 0.0265\times P_T - 2.0\times 10^{-5}\times P_T^2 + 4.484\times 10^{-9}\times P_T^3$ $- 1.2291\times E_P + 2.01\times 10^{-3}\times E_P^2 - 1.44\times 10^{-6}\times E_P^3 + 3.83\times 10^{-10}\times E_P^4$ $+ 10.8506\times D_P - 0.0886\times D_P^2 + 3.19\times 10^{-4}\times D_P^3 - 4.27\times 10^{-7}\times D_P^4$ $+ 7.1135\times N - 34.3356\times N^2 + 50.1695\times N^3 - 23.0366\times N^4$												
Productivity gradients:	32	4764	4149	82.1	17.1	0.2	74.6	23.8	28.4	0.822	39.9	14
$\eta = +1231.1 - 51.1768\times S_A + 545.1\times S_A^2 - 2096.1\times S_A^3 + 2607.4\times S_A^4$ $+ 930.2\times GDD_0 - 1873.7\times GDD_0^2 + 1169.6\times GDD_0^3$ $+ 120.4\times SDD_{10} - 1571.3\times SDD_{10}^2 + 5645.2\times SDD_{10}^3 - 6773.1\times SDD_{10}^4$ $+ - 404.7\times R_q + 28.2128\times R_q^2 - 0.6566\times R_q^3$ $+ 284.9\times V_{PD} - 363.4\times V_{PD}^2 + 157.5\times V_{PD}^3$ $+ 0.0552\times P_T - 6.0\times 10^{-5}\times P_T^2 + 2.484\times 10^{-8}\times P_T^3 - 3.82\times 10^{-12}\times P_T^4$ $+ 8.22\times 10^{-3}\times E_P$ $+ 9.9682\times D_P - 0.0797\times D_P^2 + 2.82\times 10^{-4}\times D_P^3 - 3.73\times 10^{-7}\times D_P^4$ $+ 4.1809\times N - 23.4430\times N^2 + 36.0017\times N^3 - 17.0344\times N^4$												
Thermal sum indices:	30	4744	4122	82.1	17.8	0.2	74.5	24.3	27.9	0.821	39.8	32*
Eucalyptus amygdalina:												
$\eta = -570.949 - 6.34727\times S_A + 75.4696\times S_A^2 - 150.975\times S_A^3$ $- 4.66377\times T_{mn} + 0.95530\times T_{mn}^2 - 0.074984\times T_{min}^3$ $- 165.873\times T_{max} + 24.7426\times T_{max}^2 - 1.48443\times T_{max}^3 + 0.031324\times T_{max}^4$ $- 127.97\times 10^2\times R_{Tmax} + 194.54\times 10^3\times R_{Tmax}^2 - 986.969\times 10^3\times R_{Tmax}^3$ $- 1583.84\times 10^2\times R_{Tmin} + 5685.689\times 10^3\times R_{Tmin}^2 - 8855.1332\times 10^4\times R_{Tmin}^3 + 5073.7687\times 10^5\times R_{Tmin}^4$ $+ 627.189\times R_q - 46.4776\times R_q^2 + 1.14156\times R_q^3$ $+ 1.5058\times 10^{-2}\times P_T - 1.2749\times 10^{-6}\times P_T^2 - 9.2627\times 10^{-9}\times P_T^3 + 3.5291\times 10^{-12}\times P_T^4$ $- 4.70206\times E_P + 7.5899\times 10^{-3}\times E_P^2 - 5.3481\times 10^{-6}\times E_P^3 + 1.3966\times 10^{-9}\times E_P^4$ $+ 24.3299\times D_P - 0.20623\times D_P^2 + 7.6740\times 10^{-4}\times D_P^3 - 1.0560\times 10^{-6}\times D_P^4$ $- 15.0820\times N + 49.9049\times N^2 - 55.9229\times N^3 + 18.3373\times N^4$												
Environmental gradients:	36	5291	4424	87.1	12.8	0.1	79.2	28.7	17.8	0.872	48.8	66*
$\eta = -9192.8 - 4.1326\times S_A + 64.6467\times S_A^2 - 133.6\times S_A^3$ $+ 4220.8\times A_x - 6058.5\times A_x^2 + 2921.6\times A_x^3$ $+ 80029.3\times C_{pn} - 224444\times C_{pn}^2 + 268904\times C_{pn}^3 - 117299\times C_{pn}^4$ $- 441.3\times G_c + 9.5321\times G_c^2 + 2.7057\times G_c^3 - 0.1328\times G_c^4$ $+ 777.3\times V_{PD} - 759.0\times V_{PD}^2 + 249.4\times V_{PD}^3$ $+ 0.0184\times P_T - 3.07\times 10^{-6}\times P_T^2 - 9.31\times 10^{-9}\times P_T^3 + 3.76\times 10^{-12}\times P_T^4$ $- 5.2580\times E_P + 8.52\times 10^{-3}\times E_P^2 - 6.01\times 10^{-6}\times E_P^3 + 1.565\times 10^{-9}\times E_P^4$ $+ 22.7167\times D_P - 0.1935\times D_P^2 + 7.22\times 10^{-4}\times D_P^3 - 9.97\times 10^{-7}\times D_P^4$ $- 14.5738\times N + 47.9712\times N^2 - 53.7233\times N^3 + 17.7530\times N^4$												
Productivity gradients:	33	5190	4318	86.8	13.0	0.1	79.6	27.9	17.5	0.869	48.1	120*

Table 6.5 Continued.

Model	df	Δ *Dev*	score	conc	disc	tied	%correct	F+	F–	c	% $\widetilde{R}^2$	χ^2
Eucalyptus amygdalina: continued.												
$\eta = -57.4685 + 17.3781\times S_A - 34.2302\times S_A^2$ $- 1278.5\times GDD_0 + 884.2\times GDD_0^2$ $- 551.7\times SDD_{10} + 2282.6\times SDD_{10}^2 - 6496.4\times SDD_{10}^3 + 6383.2\times SDD_{10}^4$ $- 20.4672\times R_q + 0.7418\times R_q^2$ $+ 1133.4\times V_{PD} - 1115.4\times V_{PD}^2 + 392.6\times V_{PD}^3$ $+ 0.0330\times P_T - 3.0\times10^{-5}\times P_T^2 + 7.413\times10^{-9}\times P_T^3$ $- 3.7742\times E_P + 6.28\times10^{-3}\times E_P^2 - 4.56\times10^{-6}\times E_P^3 + 1.22\times10^{-9}\times E_P^4$ $+ 24.5546\times D_P - 0.2052\times D_P^2 + 7.51\times10^{-4}\times D_P^3 - 1.02\times10^{-6}\times D_P^4$ $- 15.5355\times N + 52.7091\times N^2 - 60.1985\times N^3 + 20.4283\times N^4$												
Thermal sum indices:	28	5099	4293	86.3	13.5	0.1	78.7	29.5	18.2	0.864	47.4	79*
Eucalyptus globulus:												
$\eta = +2124.00 + 4.87732\times S_A$ $+ 0.0\times T_{min} - 20.9974\times T_{min}^2 + 2.17924\times T_{min}^3 - 0.081222\times T_{min}^4$ $+ 98.9136\times T_{max} - 0.38829\times T_{max}^2$ $- 86.6798\times T_{range}$ $- 2038.403\times10^2\times R_{Tmax} + 5253.1548\times10^3\times R_{Tmax}^2 - 5967.3248\times10^4\times R_{Tmax}^3 + 2522.3326\times10^5\times R_{Tmax}^4$ $+ 79.2560\times R_q - 2.83414\times R_q^2$ $+ 0.22817\times P_T - 3.4789\times10^{-4}\times P_T^2 + 2.2969\times10^{-7}\times P_T^3 - 5.6223\times10^{-11}\times P_T^4$ $- 0.32222\times D_P + 1.2359\times10^{-3}\times D_P^2$												
Environmental gradients:	19	655	502	74.0	25.7	0.3	78.7	41.4	20.9	0.741	18.7	11
$\eta = +5.0859 + \times S_A + \times S_A^2 + \times S_A^3 + \times S_A^4\ 79$ $+ 1393.6\times A_x - 850.5\times A_x^2$ $- 7039.2\times C_{pn} + 5101.9\times C_{pn}^2$ $+ 195.6\times G_c - 6.0525\times G_c^2$ $+ 8.8138\times V_{PD}$ $+ 0.2682\times P_T - 4.0\times10^{-4}\times P_T^2 + 2.636\times10^{-7}\times P_T^3 + - 6.36\times10^{-11}\times P_T^4$ $- 0.1145\times D_P + 3.86\times10^{-4}\times D_P^2$												
Productivity gradients:	14	592	431	72.9	26.7	0.4	78.0	78.6	21.5	0.731	17.1	8
$\eta = -1919.2 + 4.7141\times S_A$ $- 857.4\times GDD_0 + 537.4\times GDD_0^2$ $- 210.7\times SDD_{10}$ $+ 80.3745\times R_q - 2.8497\times R_q^2 + \times R_q^3$ $+ 7252.8\times V_{PD} - 12691.1\times V_{PD}^2 + 10009.8\times V_{PD}^3 - 2973.2\times V_{PD}^4$ $+ 0.2638\times P_T - 4.0\times10^{-4}\times P_T^2 + 2.641\times10^{-7}\times P_T^3 - 6.49\times10^{-11}\times P_T^4$ $- 0.0950\times E_P + 5.5\times10^{-5}\times E_P^2$ $+ 1.1785\times D_P - 7.16\times10^{-3}\times D_P^2 + 1.5\times10^{-5}\times D_P^3$												
Thermal sum indices:	19	650	492	74.1	25.5	0.3	78.2	56.0	21.1	0.743	18.6	23*

The two indices for canopy photosynthesis (light-limited and light-saturated productivity) probably contribute complementary information about the light and temperature regime of a site, but the index for canopy carbon gain that includes the effects of respiration has an opposing influence, except for the case of *E. globulus*. The typical inclusion of all productivity variables makes the interpretation of species' responses no less complex than is the case with the original direct environmental gradients for temperature and light.

The thermal indices for growing degree-days and stress-degree days were also both included in each model, with reinforcing effects except for *E. regnans*. Models with a larger number of variables generally gave a larger change in deviance, and corresponding fit statistics. Although the generalised R^2 ($\widetilde{R}^2$) values indicated that the models for *E. amygdalina* were more completely defined than for other species (up to 49%), the Hosmer and Lemeshow goodness-of-fit statistic suggests that the fitted model is not an adequate model ($p < 0.01$). Only the fitted

models of environmental gradients and productivity indices for *E. obliqua* and *E. globulus* were adequate ($p > 0.05$).

In most cases, the overall assessment of the model fits indicated that the models defined by the environmental gradients for light and temperature, in the presence of water and nutrient gradients, gave the best fits. The marginal exception was for *E. regnans*, with the best model being defined by the set of productivity indices. This may reflect the fact that this species tends to occur on the higher productivity sites (e.g. Ashton 1981).

Interpretation of these multivariate models requires that species' responses be viewed by one or two gradients. The predicted ecological responses to environmental gradients, from the multivariate models in Table 6.5 are demonstrated for the two temperature gradients, mean annual minimum and maximum temperature in Figure 6.11. The shape of the predicted ecological responses in relation to a single environmental gradient are influenced by the presence of other factors that are themselves interdependent (e.g. see correlation matrix in Table 6.2). The relative multivariate effects of the cloud of predicted responses can be demonstrated by the outer envelope and percentile of the estimated probabilities (Fig. 6.11). Each response has an outer envelope defining optimum conditions for a given set of environments within which the response declines with the limiting effect of other factors at each level of the gradient displayed.

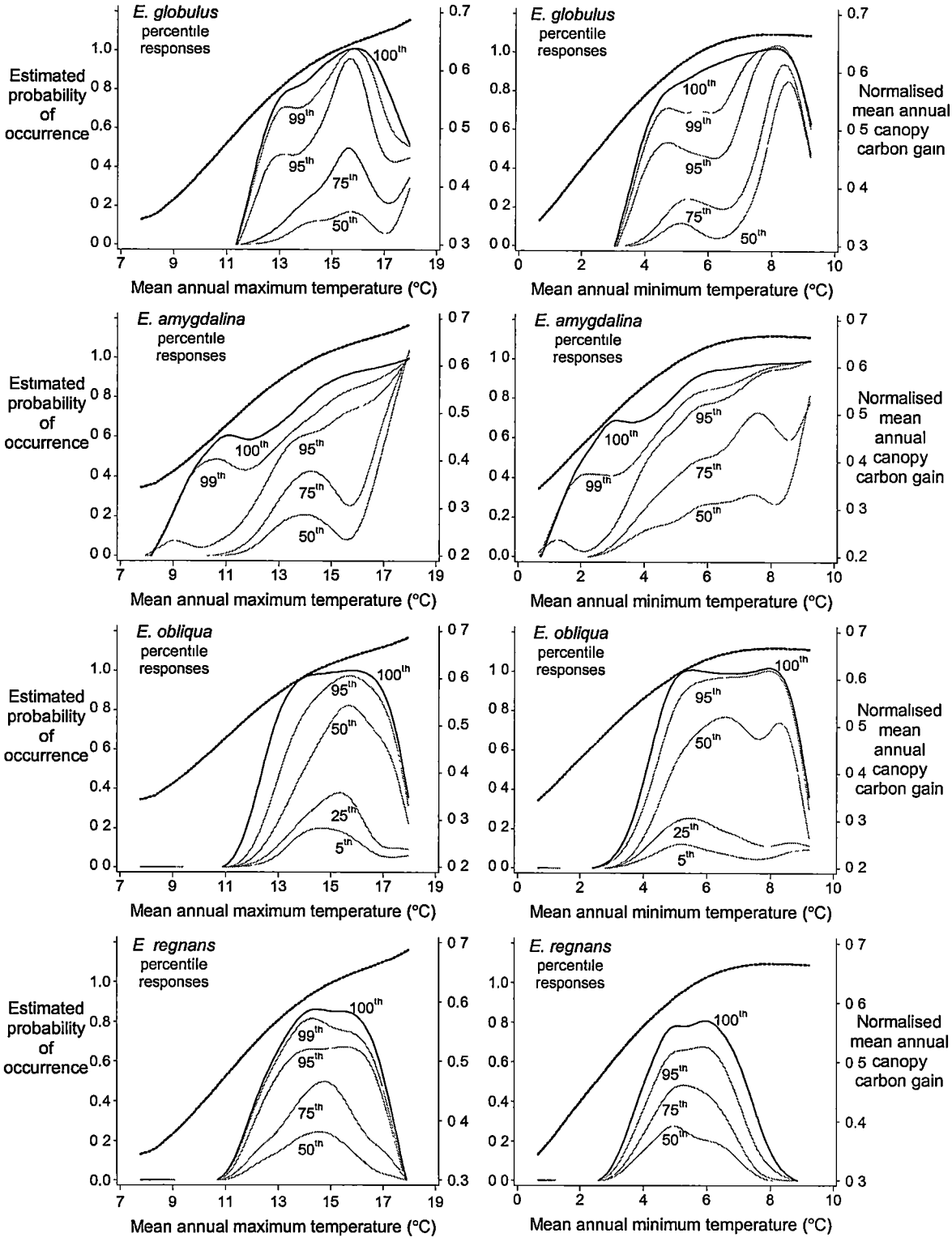


Figure 6.11 Gradient comparisons between ecological and physiological responses with variation in temperature. The predicted ecological response of each species to environmental gradients in temperature, water, light and nutrients (from Table 6.5) is presented as the outer envelope (100^{th} percentile) and other percentiles of the estimated probabilities (e.g. 99^{th} to 50^{th}) for gradients in mean annual minimum and maximum temperatures. The estimated generic physiological response for *Eucalyptus* species in Tasmania is presented as the outer envelope (n = 15 640) of the productivity gradient for mean annual canopy carbon gain (thick dotted line). Axes for each variable were scaled to facilitate the comparison of fundamental and realised niches.

The overlay of these figures with the productivity gradient for mean annual carbon gain demonstrates the potential for interpreting the relative responses in terms of fundamental and realised niches. The productivity gradient represents a generalised physiological response for *Eucalyptus* species in Tasmania, against which the ecological response of each species can be directly compared. Such comparison of ecological and physiological responses along the same environmental gradients may provide a mechanism for understanding the nature of the differences between the fundamental and realised niches of different species (Fig. 6.11). However, valid comparison of each physiological response would require the specification of genetic parameters for each species, rather than the generalised response developed here.

The percentiles of the estimated probabilities for the multivariate ecological responses to gradients in temperature demonstrate how plant distributions vary with the interaction between temperature and other environmental factors in the model (Fig. 6.11). For example, the predicted ecological response of *E. globulus* to mean annual maximum temperature demonstrates a persistent optimum at about 16°C as other environmental factors increasingly limit the response. The optimum response to mean annual minimum temperature is about 8°C. However, when all other environmental factors are optimal, the outer envelope of the predicted ecological response suggests that a broader tolerance of cooler temperatures would promote the occurrence of *E. globulus* across a wider geographic range. Similar interpretation of the ecological responses of the other species is also possible. For example, *E. amygdalina* extends onto cooler sites when all other environmental factors are optimal, but this temperature response declines dramatically with other factors limiting. The limiting minimum and maximum temperatures shift from about 1°C and 8°C to about 2.5°C and 10°C for the 95th percentiles of the estimated probabilities. The complex response suggests that more than one ecotype may be involved in the distribution of *E amygdalina*, depending upon the limiting conditions of cold or drought stress that contrast the ends of its environmental range.

The multivariate ecological responses are more clearly defined for the distributions of *E. obliqua* and *E. regnans*. *Eucalyptus obliqua* demonstrates a broader tolerance of minimum temperatures when all other conditions are optimal, than does *E. regnans* (5-8°C versus 4-6°C). For the 50th percentile of estimated probabilities, the optimum responses to minimum and maximum temperatures for these two species show a clear separation (6.5°C and 15.5°C for *E. obliqua* versus 5°C and 14.5°C for *E. regnans*). However, *E. obliqua* may sparsely co-occur with *E. regnans* on sites where *E. regnans* is expected to dominate the canopy (e.g. compare optimal temperatures for the 25th percentile of *E. obliqua* with the 50th percentile of *E. regnans*).

The simple correlation between the predicted ecological responses (Fig. 6.11) and the generalised physiological response for eucalypts of mean annual carbon gain indicates the position of each species relative to a gradient in productivity (Fig. 6.12). These responses demonstrate the usefulness of a standardisation of environmental factors in terms of plant growth characteristics for exploratory investigations of the nature of ecological interactions between species. All

species shown tend toward an optimum response on sites of higher potential productivity (> 0.5, normalised mean annual canopy carbon gain). The predicted optimum ecological responses (outer envelopes) for *E. obliqua* and *E. regnans* appear as nested responses. *Eucalyptus amygdalina* also completely overlaps with *E. globulus* for its optimum response, but also extends onto sites of much lower productivity than any of the other three species. The complex response of *E. amygdalina* probably reflects clines in genetic characteristics within its species' range or intergrades with *E. coccifera* and *E. nitida* in cooler highland and western region sites (e.g. see discussion in Williams & Potts 1996). This latter response is reflected in the broader confidence intervals which are associated with these predicted responses for *E. amygdalina* on cooler sites.

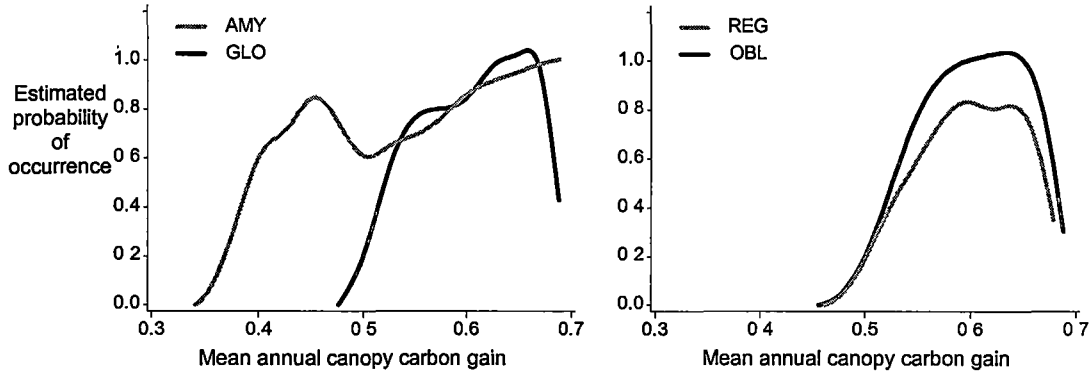


Figure 6.12 Outer envelope of predicted ecological responses of four *Eucalyptus* species by mean annual carbon gain. Ecological responses are defined by multivariate models of environmental gradients (Table 6.5), and are graphically presented relative to a productivity gradient defined from physiological parameters for photosynthesis and respiration.

6.4 Discussion

The analyses of plant responses to direct gradients in light and temperature and their derivatives as thermal sum or productivity indices indicate that different indices were better for different species, but that some indices were adequate for all species. The potential significance of more than one type of thermal index suggests that different species' responses may be distinguished by competing effects of growth or stress conditions in their habitats. Critical temperature values for the thermal sum indices (growing degree days and stress degree days), however, assume all species have the same general response. An alternative approach to the derivation of these indices, in the absence of specific information about the physiology of a species may be iterative testing. For example, physiological differences between species could be inferred by iteratively testing the critical temperature values for the thermal sum indices which result in the best fits to a species' distribution. A similar approach to modelling species' distribution was undertaken by Sykes *et al.* (1996).

A standardised vegetation productivity gradient was expected to improve the modelled responses and increase the ecological interpretability of species' comparisons (Fig. 6.9). However, the strong correlation between minimum and maximum temperature variables and growth or stress conditions, suggests that a fairly direct interpretation of response patterns can also be obtained

from these simpler variables. Nevertheless, the index for normalised canopy photosynthetic productivity simplifies the interpretation of species' responses compared with mean annual maximum temperature. For example, the responses of *E. globulus* and *E. amygdalina* were offset, and the separation of the optimum responses between *E. regnans* and *E. obliqua* were clarified by the skewed shape from a cubic polynomial for normalised canopy photosynthetic productivity (Fig. 6.9).

The bivariate response of *E. globulus* and *E. amygdalina* to environmental variation in annual and seasonal temperature indicates that warmer sites were preferred and that a wider range in diurnal temperatures can be tolerated on such sites. However, on cooler sites, *E. globulus* occupies a narrower range of diurnal temperature variation, but *E. amygdalina* extends across a wider range of diurnal temperatures (Fig. 6.10). This is consistent with the relatively coastal occurrence of *E. globulus* in south eastern Tasmania and the more widespread distribution of *E. amygdalina* into northern and inland Tasmania (e.g. Duncan & Brown 1985; see also Williams & Potts 1996, pp. 48, 69). However, the overlapping responses between *E. amygdalina* and *E. globulus* (e.g. Fig. 6.9) do not necessarily suggest intense competition, but rather that the two species frequently co-occur (e.g. Duncan & Brown 1985). In dry sclerophyll forest, for example, *E. globulus* is commonly observed as a sparse, emergent tree above a canopy of a series *Piperitae* species. Although these generalised ecological inferences did not change, the actual variables that are of most significance were influenced by other relatively independent environmental gradients (e.g. water, nutrients and light).

Warmer sites were also preferred by *E. regnans* and *E. obliqua*, with *E. regnans* extending onto cooler sites with a wider diurnal variation in temperature than tolerated by *E. obliqua*. The thermal regimes of the habitats occupied by *E. regnans* and *E. obliqua* were clearly distinguished, consistent with the possibility of intense competition between these species where they overlap in distribution (e.g. Ashton 1981). However, the distributions of *E. regnans* and *E. globulus* were best described by minimum temperature regimes, possibly reflecting an intolerance for frost where this may be a recurrent feature of its habitat (e.g. Kirkpatrick 1975a, b; Paton 1980; Rook *et al.* 1980; Griffin *et al.* 1982; MacLeod 1981, 1982). Support for such hypotheses may also explain some regeneration problems experienced with *E. regnans* where changes in the drainage of cold air have followed extensive clearing of forest habitat in broad inland valleys of Tasmania.

In the case of *E. obliqua*, the annual thermal magnitude best describes the distribution patterns because the annual growth potential, irrespective of the season, may be more important in determining the outcome of competition (e.g. Chapter 5; West 1981; Cotterill *et al.* 1985). This trend was further emphasised by its slightly improved response to a gradient in light-saturated photosynthetic rates, over responses to the thermal sum and day-time average temperature. Whether this relationship for *E. obliqua* was real or coincidence, the response was evidence for

the potential improvement in ecological modelling from a productivity gradient defined by the appropriate genetic parameters for each species.

The examples of multivariate modelling were simplified by fitting only the mean annual estimates for each factor, rather than the additional gradients for the seasonal effects denoted by absolute minimum and maximum values. However, the univariate responses in Table 6.3 demonstrated that species were responding to these absolute values for an average season, rather than necessarily the mean annual variables. Multivariate models based on these additional factors may better reflect the average seasonal variability of climate, and therefore the limits to plant distribution.

Subsequent multivariate models may need to comprise a very large number of variables (and their polynomials to define response shapes) to adequately define the ecological responses of each species. This is because there is a large number of distinct physiological and ecological traits which need to be distinguished between species. Each environmental factor may act as a surrogate for the dominant expression of one or more of these traits. In this respect, the productivity gradients also represent only one or a few traits of a species' distribution - those most correlated with the growth response of a species. But there are many other factors that may be significant in determining a species' distribution. Regeneration characteristics may be of particular importance, and the potential role of disturbance regimes such as fire and frost, that favour one species over another at a site. The way these factors combine together may be highly influential in the outcome of some species' distributions. Plant distribution patterns are therefore ultimately determined by the interaction between the traits of different species. These become apparent as their ecological responses, from the correlation between occurrence and environment.

Generally, the productivity indices for temperature and light responses are no better or worse than the original climatic variables as predictors of plant species' distribution. While these canopy productivity gradients have potential for predictive modelling, a similar index may be needed for each physiological trait that reflects adaptive responses such as respiration, seasonal acclimation, and cues for flowering and regeneration, including the effects of water limitation. In the absence of such detail, polynomials of the individual gradients in water, temperature and light from climate may be considered as surrogates for such physiological responses. These adequately account for the species' response to the average and extreme conditions that fluctuate in importance throughout its range. In addition, the original climate variables and their empirical derivatives (e.g. vapour pressure deficit, rates of change) do not require physiological assumptions about how plants respond. However, interpretation of species' ecological responses to these gradients necessarily requires some indirect inference about the likely suite of adaptive traits and physiological performances that may have contributed to the observed distribution patterns.

The use of productivity as a surrogate for plant response to environment is contentious because plants are not expected to respond to gradients in temperature, light, water, nutrients, or carbon dioxide in the same way (Austin & Smith 1989; Austin & Gaywood 1994). However, where an index of productivity is based on known physiological processes then it may have application to the standardisation of plant performance and functional comparison of species.

Productivity gradients may therefore have application in niche studies. Ecological responses from environmental gradient analyses could be compared with the physiological responses from process models of plant performance. Hypotheses about the role of competition and other historical and disturbance factors in constraining species' distributions, relative to their potential, can then be explored. Even without comparison with physiological responses, the ecological response models defined from direct environmental gradients provide the basis for refining questions about the ecology of a species that can be more rigorously tested in experimental designs.

Oecologia (1996) 108.518–528

M. Battaglia · K.J. Williams

Mixed species stands of eucalypts as ecotones on a water supply gradient

Received 25 September 1995 / Accepted 14 April 1996